

**On how the nature of mental events
may affect their role in psychological explanation**

Submitted for PhD
by Paul Jonathan Pitt Noordhof,
University College London

ProQuest Number: 10608899

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10608899

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

An attempt is made to assess the likely role of mental events in psychological explanation. Various ways in which it might be thought that mental events are explanatorily idle are considered. Some reasons are given for singling out two arguments for special consideration.

The first argument is that the physical world is causally closed, yet mental events are non-physical, hence mental events cannot be cited by psychology in the causal explanation of behaviour. To assess it, an account of the nature of the physical is put forward, and the premises that constitute the first argument refined and discussed. The existence of phenomenal and intentional properties is stated to be the main motivation for claiming that mental events are non-physical. It is claimed that there is an argument for the non-physical character of phenomenal properties based upon an understanding of awareness. In contrast, it is held that there is no reason to suppose intentional properties are non-physical. It is suggested that, given our conclusion about phenomenal properties, there is reason to reject the claim that the physical world is causally closed.

The second argument holds that while psychology should only postulate explanatory entities whose occurrence is metaphysically independent of the environment in which the subject is located, some mental events with intentional properties are not independent, therefore, such mental events should not be cited in psychological explanations. It is argued that psychology should only take such an attitude to these mental events, if it is just concerned with the causal explanation of behaviour. An alternative view of psychological explanation is put forward involving, at one level, an appeal to norms, which allows the mental events mentioned an explanatory role, contrary to the original claim.

Therefore, both arguments are rejected.

C o n t e n t s

	page
Introduction	5
Chapter 1 : The first challenge of epiphenomenalism.	22
Chapter 2 : Refinement of the first argument	68
Chapter 3 : The extent of the challenge.	98
Chapter 4 : Two ways in which mental events might be non-physical.	134
Chapter 5 : On the existence of phenomenal properties.	150
Chapter 6 : An argument for the proposition that phenomenal properties are not physical.	179
Chapter 7 : On the nature of intentional properties	212
Chapter 8 : The second challenge of epiphenomenalism.	255
Chapter 9 : On the possibility of a species of explanation based on norms.	273
Chapter 10 : On the likelihood of integration of norm-based explanation into psychology.	295
Conclusion	319
Bibliography	333

Introduction

The aim of this work is to examine the likely role that mental events will play in psychological explanation. More specifically, we will be concerned with whether the nature of mental events throws into question their role in such explanation. We will be interested in the nature of mental events, but only in so far as an understanding of this nature will help us in this aim. The phrase 'Mental events' will be used throughout the present work to stand for a category which includes any of the following, beliefs, desires, judgements, thoughts, sensations, perceptions, moods, hopes, fears, intentions, imaginings, dreamings and so on, for those occasions when that level of generality is appropriate. In fact, 'mental events' will be understood so broadly as to include any of the entities that have been traditionally supposed to be mental phenomena.

Sometimes the use of 'mental event' will seem unhappy, no doubt especially when it occurs in a way which purportedly clashes with a common conception of some type of mental entity. For example, it is probably felt that a sentence starting 'Belief is a mental event ...' does not express the thought we shall suppose it to, rather better would be the sentence starting 'Belief is a mental state ...'. It may well be claimed that the first sentence must be false whereas the second sentence may be

true, depending, of course, upon how we complete it. However, once the stipulation we have made is remembered, no opposition should arise.

The limitation of the phrase 'mental events' to cover just those mental phenomena traditionally identified is not meant to commit us to the thesis that we already know the types of mental phenomena there are and need not fear the discovery of new ones. Such discoveries may well be possible. The limitation is only in the cause of abbreviation.

As has already been indicated, the question with which the present work is concerned is this: Does the nature of mental events in any way throw into question their role in psychological explanation? There is something rather puzzling about this question and it is the nature of the puzzlement that we should be clear about before we proceed further. In sketching out the difficulty, it is to be hoped the motivation for structuring the present work, in the fashion it has been, will be brought into view.

Any question concerning the nature of mental events may be thought to be the subject matter of psychology, and, if this is true, it may seem somewhat surprising that we are being faced with a question about how the nature of mental events may affect the role they play in psychological explanation. Instead, it may be thought

that mental events are part of what needs to be explained and not part of any explanation. They occur as the 'explanandum' and not the 'explanans'.

Although mental events are part of what needs to be explained, they are also quite legitimately considered to be a possible part of what a psychological explanation refers to in order to explain some other phenomena that it is psychology's task to explain. The most important examples of this are the following: first, the explanation of someone's behaviour by reference to the mental events he or she had which preceded it; second, the explanation of the occurrence of some mental events by citing mental events which preceded them; third, the explanation of the nature of complex mental events from simple mental events. The last type of explanation would include, for example, providing an account of the nature of practical deliberation, a complex mental event, by citing the beliefs and desires which constituted it and which resulted in a particular decision to act in a certain way. So, it is reasonable to suppose that mental events may have a role in psychological explanation. The question is whether they are likely to have such a role.

A second and connected question arises with respect to our inquiry. It is over whether one should attempt to discuss the nature of mental events in a philosophical

work. After all, whatever philosophy is, it is not a discipline which prides itself on being centred on empirical research. Yet, if we are to find out the nature of a certain sort of element which exists in the world, namely mental events, empirical research seems exactly what is required.

Part of the reason why a philosophical approach is relevant is that the answer to this question is not one which seems to be attainable just through obtaining some more data. We need to look more carefully at the data we do have, and philosophy can help in this respect. Another reason is that the answer involves coming to a preliminary view about the nature of explanation, and psychological explanation in particular, which should be reached prior to psychological investigation rather than consequent on it. There could be no coherent psychological investigation without having some view about the type of explanation for which such investigation is searching. Of course, the latter point is not to rule out a change in our conception of psychological explanation, as psychological research proceeds. A final reason for the philosophical nature of our inquiry is that some philosophers have spoken on the question which is our concern, and offered philosophical arguments to establish conclusions in this area. Consequently, we may suppose that we are answering certain inquirers in their own

coinage. 'When in Rome

Although there is more to be said on this issue, to do so would be to take us too far afield. It is to be hoped that the content of our discussion will demonstrate the relevance of philosophy to the issue.

The subsequent inquiry will be concerned primarily with two ways in which the nature of mental events may effect their role in psychological explanation. The orientation of our attempt to answer the general question will be defended in Chapter 3. There we will argue that the other ways, in which, it has been claimed, the explanatory role of mental events is threatened by their nature, are not as convincing. In particular, we will hold that there is no obvious inference to make from the claim that the mental supervenes on the physical to the claim that its role in psychological explanation is open to question. The reason for the justification of our approach being postponed until then is not natural timidity, but rather its reliance upon some of the discussion in the second chapter.

The first way we shall consider at length, in which the supposed nature of mental events may effect their role in psychological explanation, stems from an acceptance of the following principle, concerning the physical realm, a principle which has seemed intuitive to many:

The physical realm is a closed system, in other words, nothing non-physical can have a causal relationship with anything physical.

Now it is uncontroversially true that, under any normal understanding of what is physical, the behaviour of human subjects is physical. It is also not open to serious question that psychology amongst other things, is interested in the explanation of human behaviour. On the assumption that

to explain human behaviour you need to cite at least some of its causal antecedents, it looks as if the explanatory role of mental events may be threatened. It has seemed plausible to many that mental events are non-physical. From which it would seem to follow that mental events could never figure in an explanation of someone's behaviour. But, as we noted, psychological explanation is committed to the explanation of human behaviour. Therefore, it would appear, mental events have no role in such explanations. One of the philosophical tasks of the present work, will be to assess the argument just given.

How may we do this? The first thing we ought to do is come to an understanding of the nature of the physical. Only then will we have some idea of what it would be for something to be non-physical. After setting out the argument we have just offered more formally at the

beginning of Chapter 1, this is the first question that we shall address. The rest of the chapter will be devoted to it. There, we shall argue that one account of the physical contrasts favourably with all others that have been offered. Put crudely, it holds that

an entity is physical if and only if, first, it is spatio-temporarily located, second, its existence does not imply that a subject is aware of it.

There will be some refinements to this formulation later. It is cast in terms of the notion of an entity because it is meant to be applicable to all manner of metaphysical categories: substances, properties, events and so on. It is this account of the physical which is adopted.

A preliminary defence of the principle concerning the causal closure of the physical realm is offered in Chapter 1, however the bulk of its defence is contained within Chapter 2, requiring, as it did, the understanding of the physical just described. The need for some sort of 'closure principle' is argued for, in Chapter 1, on the grounds that only if scientific inquiry presumes that it is true, will it follow both that there is some reason internal to the inquiry for believing in the truth of general statements or laws and that there is a limit placed on the types of explanations that may be offered. In Chapter 2, it is held that the closure principle should concern physical entities. Two arguments are offered for

this view. A 'methodological' argument to the effect that only physical entities are explanatorily legitimate, and an 'empirical' argument to the effect that only citation of physical entities in explanations is empirically plausible. As a consequence of these arguments, a refined version of the principle is provisionally accepted.

Clearly, the soundness of the argument then rests upon the claims that behaviour is physical, some mental events are non-physical, and an explanation of behaviour involves the citation of a cause. It is assumed for the purposes of discussion that the claims concerning the nature and explanation of behaviour are correct. Therefore, the discussion focuses upon whether some mental events are non-physical.

Two properties that some hold are constituent parts of mental events are identified in Chapter 4, phenomenal and intentional properties. It is said that should these properties exist, they offer a reasonable subject matter for our discussion. Concerning both, we have the intuition that they have causal efficacy, yet also, we have the intuition that they are non-physical. These intuitions appear in conflict and a meditation upon this conflict is likely to make us appreciate whether or not the first argument is sound.

A positive argument for the claim that phenomenal properties exist is put forward in Chapter 5. It is said

that there is a prima facie case for their existence given that we putatively experience them, can talk about them, and, understand talk about them. The prima facie case can only be undermined if there are strong arguments against their existence. Three such candidate arguments are considered and rejected for various reasons.

Having defended the claim that phenomenal properties exist, the attention turns in Chapter 6 to whether they are physical or non-physical. In the first part of that chapter, it is claimed that serious difficulties face any attempt to demonstrate that some mental property is non-physical, on the basis of our experience of its nature. Thus, it seems possible that phenomenal properties are physical. However, in the second part of the chapter, it is held that there is an argument which avoids the difficulties identified and which establishes that phenomenal properties are non-physical. It rests upon the thesis that the only way in which a subject's awareness of his or her own mental events can be explained is by supposing that certain properties, the phenomenal properties, have a nature that implies that a subject is aware of them. So, we have a positive result for one of our two properties.

Turning then to intentional properties, Chapter 7 begins by making what is, hopefully, an uncontroversial claim, namely, that,

any account of intentional properties must allow that the subjects to whom they are ascribed can make various kinds of errors of judgement.

Yet, most accounts that characterise intentional properties in physical terms, cannot fulfil this simple requirement upon the attribution of intentional properties. According to each of these accounts, the subjects cannot make the various types of mistakes that are identified. This time, though, the argument of the chapter is not presumed to establish that intentional properties are non-physical, by our definition. It is not clear that allowing intentional properties to be non-physical helps matters. Rather, it is suggested that some motivation is supplied to adopt the approach to intentional properties that is introduced at the end of Chapter 7. This approach may enable us to conclude that intentional properties are physical, but, unfortunately, it raises a problem that it shares with some of the other accounts we considered, namely, the second threat to the explanatory role of mental events in psychology, to which we shall now turn.

The second threat derives from the following principle that is said to constrain the sort of entities which a psychologist will refer to in the explanations he or she offers:

All entities proposed by psychology to be explanatory, of the range of phenomena with which it is concerned, may be parts of the (psychological) subject, or person, independent of the way the subject or person's world is.

Three points about this principle need to be noted briefly. First, here, and throughout the present work, 'subject' and 'person' will be used as stylistic variants for 'creature with a mind'. Second, the sense of 'independent' appealed to in the principle is that of metaphysical independence. Crudely, one thing is metaphysically independent of something else if it is logically or metaphysically possible that the first thing exists without the second existing. The relationship of metaphysical independence is asymmetric. If one thing is metaphysically independent of something else, it does not follow that the latter is metaphysically independent of the former. Consequently, when it is asserted that the entities proposed by psychology to be explanatory should be, in this sense, independent of the way the world is, it does not follow that the way the world is may not entail that a subject is undergoing certain experiences, say. For all that the principle says, the existence of a hedgehog suitably situated may entail that we have an experience as of a hedgehog. What the principle rules out is that one could not have such an experience without the

hedgehog being so situated. The reason for stating the principle in terms of metaphysical independence as opposed to some other sort will become apparent in due course. As to our lack of commitment over whether metaphysical independence should be understood in terms of logical or metaphysical possibility, that is quite deliberate. The relationship between these two supposedly distinct sorts of modality is controversial. Nothing that follows, in our discussion of this argument, requires us to take sides on this issue, so we shall remain neutral.

There are many reasons for finding the principle plausible. We shall just touch on them here. Psychology is primarily concerned with the workings of minds and the consequences of such workings. It is to be admitted that the way the world is often has an affect upon our mental lives. It is quite compatible with this influence to believe that the actual having of a psychologically significant mental life is a matter of the person only and not the environment being in a certain state. It is equally compatible to hold that while the way the world is may affect how we behave, it only does so when mediated, so to speak, by the mind. Put these two thoughts together and we may get the idea that psychology is interested in the causal efficacy of events beneath the person's skin, as these events give rise to other mental events and behaviour. Naturally there are qualifications to be made,

but we shall save them for now.

Acceptance of the principle just specified can once more seem to threaten the role of mental events in psychological explanation. This time, the threat stems from a series of philosophical arguments which seek to illuminate the nature of mental events by highlighting one aspect of our conception of some of them, namely that:

Some mental events are dependent upon the environment, although not necessarily the present one, for their existence.

Once more, the notion of dependence being appealed to is that of metaphysical dependence, the converse of metaphysical independence. If the claim just made is true, it would seem to follow in a quite straightforward manner that these mental events will have no role to play in psychological explanation.

In trying to assess the likelihood of some mental events failing to play a role in psychological explanation in this second case, it would be preferable if we could consider the arguments that are offered for the 'environment-dependence', as we shall call it, of these mental events, and the variations that exist over the interpretation of this thesis. Questions of space, however, rule this approach out. That does not mean that we shall leave the premiss concerning the dependence of some mental events on the environment entirely

unsupported. For, we are fortunate enough, or perhaps unfortunate enough, to have indirect evidence for its truth. The account of intentional properties that we said does satisfy the condition we identified at the beginning of Chapter 7 renders some intentional properties environment-dependent. Since intentional properties are the constituent properties of some mental events a consequence of this is that some mental events are environment-dependent.

In Chapter 8 we tentatively endorse an argument for the conclusion that the environment-dependent aspects of mental events have no distinctive causal powers. Further, it is said, if one then assumes that psychological explanation is just causal explanation, it follows that these environment-dependent aspects have no explanatory role in psychology. This is the main motivation behind the principle we described a little earlier as a constraint upon the type of entities identified in psychological explanation.

Rather than accepting the argument just sketched, the principle just mentioned is rejected. Its rejection depends upon two things. First, an alternative explanatory role for the environment-dependent aspect of mental events must be provided. In Chapter 9, a style of explanation is put forward, called Norm-based explanation, which is said to supply precisely this role. The

environment-dependent aspect of mental events have a role to play in our understanding of how a mind is aimed at the truth, or, so it is suggested. Various objections to there being a style of explanation of this sort are discussed and rejected.

The second thing upon which our rejection of the argument sketched above, depends, is the claim that the style of explanation identified is a genuine case of psychological explanation, rather than some other sort. It is to this issue that Chapter 10 is devoted. There it is argued that norm-based explanation of the sort sketched in Chapter 9 is not significantly different from other types of explanation that have been put forward in cognitive science.

The position at which we will arrive is that there is no problem about the explanatory role of mental events as far as intentional properties are concerned, but there potentially is with regard to phenomenal properties. However, in the conclusion, it is argued that the fact of phenomenal properties being non-physical in itself is a reason for rejecting the principle concerning the causal closure of the physical realm. Not a reason because we do not like the upshot of the argument, but rather, because the non-physical character of phenomenal properties raised methodological and scientific difficulties more pressing than those that the adoption of the principle was

taken to ameliorate. In particular, it is argued there must be some reason for the occurrence of phenomenal properties along with physical properties which itself implies that there is some causal relationship between them. Hence, the principle is false as it stands. A reformulation of the principle is offered which depends upon their corresponding to our notion of causal priority some 'objective reality'. It is stated that it is hard to justify such an approach, but, if it is justified there are still other more plausible theses to hold than that phenomenal properties have no causal efficacy, and perhaps, thereby, no role in psychological explanation.

An important presupposition of our inquiry should be mentioned before we begin in earnest. The arguments we are considering are generally thought to have rebarbative conclusions. Moreover, there are some consequences of these conclusions which are even more unattractive, if that is possible. By mentioning these, it would have been possible to undermine faith in the argument in other ways. However, we have chosen not to do this. It is assumed that if we find the conclusion of the arguments unacceptable then one of the premises must be false. It is not that the world is incoherent, or that we have a radically and unrevisably incoherent grasp of it. It is further assumed that if we must reject one of the premises of the arguments, then we should be able to find a reason

to do so independently of the unattractiveness of the conclusion. Some might view this presupposition as overly optimistic in this 'post-modern age' - but there we go.

Perhaps the order of our investigation will seem surprising. It may be thought that the neatest way to approach the problems described would be to come to an understanding of psychological explanation first and use this to determine the likely role of certain mental events, in such explanation, in subsequent chapters. The reason why this way of approaching the subject was not adopted is that there is no reason to suppose that we have any clear idea of what we expect from psychological explanation, and bearing this in mind, it would serve us well not to prejudge this matter right at the beginning. It is undeniable that mental events, of the sort we have indicated, have been used to give primitive psychological explanations. If we now wish to conclude that the mental events in question are not up to scratch, we would be well advised to first make sure that we have not overlooked some explanatory role that they play, a role which is revealed by their nature, but to which we are blinded, because we expected psychology to match the experience we perhaps take ourselves to have had in other sciences. It is to be hoped that the way that has been chosen to approach the subject matter illuminates more than it obscures.

Chapter 1

The first challenge of epiphenomenalism.

Our question is 'Does the nature of mental events in any way throw into doubt their role in psychological explanation?' One of the ways in which the nature of mental events may raise exactly this doubt is if it is claimed that they are not physical. The reason for this will be examined in the first part of the chapter. There we shall outline an argument that purportedly demonstrates that if the mental is non-physical, it has no role to play in psychological explanation. Supposing the argument to be valid, it will be of interest to us to determine whether mental events are non-physical and, to do this, it would be helpful if we had some understanding of the nature of the physical. Only with such an understanding would we then be in a position to assess whether mental events are non-physical. The second part of this chapter will present constraints upon any account of the physical. These constraints will be important for our assessment of the various theories put forward concerning the nature of the physical that will occupy the third part of the chapter. The discussion in this part is supposed to bring out some issues which will be useful in motivating our own account of the physical, offered in the final part.

(a) An initial formulation and discussion of the argument

The argument we are going to consider has already been sketched in the introduction. We shall try to assess its force in the chapters that follow. First, we shall state it in summary form.

- (1) The physical realm is a causally closed system; in other words, nothing non-physical can have a causal relationship with anything physical.
- (2) All mental events are non-physical.
- (3) Human behaviour is physical.
- (4) If mental events are to have a role in the psychological explanation of behaviour, then they are among the causes of behaviour.
- (5) No mental events can have a causal relationship with human behaviour.

Therefore:

- (6) It is not the case that mental events have a role in the psychological explanation of human behaviour.

The conclusion seems to commit those who accept the premises to a limited form of 'epiphenomenalism' which, for the moment, we shall take to be the thesis that, in this case, mental entities have no affect upon other entities, specifically behaviour. In the present section, we will discuss what is surely the hardest premiss to

understand, premiss (1). The full argument for it must await the discussion of Chapter 2, but we are in a position to appreciate some of the motivation behind it, now.

The claim that the physical realm is causally closed has seemed intuitive to many.¹ While it no doubt cannot be demonstrated to be true, it can perhaps be shown that it must be presumed to be true for scientific inquiry to proceed. This is what we shall try to do. It is important to be clear at the outset what is meant by 'scientific inquiry' here. It does not refer to the inquiry of any particular science. Rather, it refers to that programme of investigation upon which the sciences are jointly embarked. So, the claims that follow should not be taken to be true of particular sciences.

One of the objects of scientific inquiry is the discovery of laws, or, at worst, general statements concerning the succession of events, if laws are considered inappropriate or unavailable with regard to a particular domain of inquiry. We shall call them laws from now on, but bear in mind this qualification. One example of a law concerning events is 'After every clap of thunder, there is a flash of lightning'. For all I know this may be false, but that does not undermine the illustration. The laws that are put forward during scientific inquiry can only merit our confidence under

certain circumstances. It is by examining the nature of those circumstances in which we feel confident in putting forward laws that we shall see the motivation for claiming that the physical realm is a causally closed system. First, let us become clear as to the nature of systems, and causally closed systems in particular.

Generally speaking, a system may be defined to be any collection of things around which one may draw a boundary so that members of the system are demarcated and other entities thereby excluded. The boundary need not be spatial. A causally closed system is one for which the entities specified as members of that system do not enter into any causal relations with entities which are not members of that system.

Suppose now, we notice an association between two types of events, type A and type B. If an event of type A occurs, then, we have observed, an event of type B shortly follows. Under what circumstances would we be confident in asserting, as a law, that events of type A are followed by events of type B? The immediate answer is that we would be confident only if we became convinced that the association we have observed would not fail, should the circumstances in which an event of type A occurs, be different. We want to rule out the association being a consequence of the particular circumstances in which we have observed them up until now.

It is, of course very difficult to rule out the possibility that the association we have observed is just a consequence of the circumstances in which we have observed it. What would help is if we had some idea of the sorts of entities that may interfere with the association so that on the occasion when these entities are present the association does not occur. These interfering entities it seems reasonable to call 'causal factors' as it is hard to understand the sort of interference that they provide unless it is causal. It is in the light of this that we can see that the claim that the physical realm is a causally closed system has a role to play. If we are considering an association between physical events, and we know that the physical realm is causally closed, and we have some appreciation of what a physical entity is, then the types of entities that are possible interfering factors are limited. We do not have to consider every type of entity, just physical ones.

The basic idea, then, is that we presume the physical realm is a causally closed system and in so doing simplify our task of discovering 'physical' laws, laws connecting physical entities to each other. The fact that we have a simpler task means that when we are considering whether an association between two types of event is in fact the result of a law, there is less that we must do, to have reason to believe this is so.

It is important to be clear about the status of the reason that the presumption we have described provides. There is a sense in which we can never have reason to believe that a certain type of law is true, that concerning the unobserved. David Hume² noted that our observation of an association between two types of things can never give us reason to believe that the hitherto unobserved instances of these types of things will be likewise associated unless we have reason to believe that types of things will behave in the same way whether we observe them or not, and this we cannot have. The reason that we think the existence of a closed system will provide for our belief that a certain type of law holds is not a reason which provides a reply to Hume's sceptical argument against induction. Instead, the 'reason' is only a reason conditional upon the conclusion one comes to, as a result of Hume's argument, about the rationality of scientific 'reasoning' generally. Indeed, the 'reason' is also conditional upon the rationality of adopting the claim that a certain type of system is causally closed.

The assumption that a certain system is causally closed does not just provide us with reason to believe that a certain observed association is the result of a law. It also helps us to limit the type of explanation that we should seek. Suppose, for instance, we presumed that the spatio-temporal world is causally closed. Before

our eyes, a marble statue flattens into a square block. What explanation should we offer? Well, here is one that we can rule out. The pressure of God's finger tips situated outside space and time had the effect mentioned. Instead, we search for an explanation within space and time. The claim that a certain system is causally closed is one way of ruling out certain types of explanations, often those which are thought to be supernatural, or whatever. In saying that a system is causally closed we do not have to say that 'supernatural' entities do not exist, a bold claim perhaps, we need only say that even if they do exist they would not be useful from the explanatory point of view. Of course, the ruling out of supernatural explanations depends upon the type of system that we say is causally closed. We could choose to rule them in rather than out.

The interesting question now is, what reason have we to think that the claim that the physical realm is a causally closed system will be useful for scientific inquiry? It is clearly a question we cannot answer until we arrive at a better understanding of the nature of the physical. The rest of the chapter will be concerned with developing such an understanding.

(b) Two constraints on any account of the physical
which is offered

A considerable amount of attention in philosophy has been devoted to what is known as the mind-body problem, that is, the question of the relationship between a person's mind and the body which we take them to possess. The relationship between this problem, and the notions of the mental, physical and non-physical is complex and in becoming clear about this relationship, we shall be able to advance the first constraint upon any account of the physical.

Often the questions 'What is the relationship between the mental and the physical?' and 'What is the relationship between mind and body?' are taken to be more or less identical. However, if one understands the first question to be asking for the connection between the two categories picked out by the phrases 'the mental' and 'the physical', it soon becomes apparent that this question is rather different from the second question in a respect which is important for us to recognise. The question concerning the categories of the mental and the physical will be answered by saying that these categories are mutually exclusive, or that they have members in common, or, that one is a sub-category of the other. However, the answers that we might give to the question what is the relationship between the mind and the body are indifferent

to, by being independent of, any of the answers just canvassed.

Suppose that one managed to convince oneself that one could continue to exist as a mind without any part of one's body existing. Presumably, the conviction would stem from the thought that the properties which one takes oneself to apprehend in one's mental life could continue to be even if all the properties which constituted one's body vanished from the world. No part of the means by which one comes to this conviction need appeal to the peculiar nature of the properties one takes to be revealed to one mentally. If there are two chairs standing side by side, one may believe that one could exist without the other. From that it does not follow that the chair in question is made of very different properties. Thus, the thought that mind and body are separate things has no implications for one's view about the relationship between the mental and the physical. One could view these two categories as co-extensive and yet come to the conclusion just mentioned. Equally, if one thought that mind and body could not be separated, there would be a temptation to view the mind as composed from some of the properties of the body, and in the light of this temptation it might seem that there are implications as to the relationship between the mental and the physical. However, this would be to overlook the possibility that the mind could have

properties which are not properties of the body and the existence of these other properties would leave open all the possible relationships between the mental and the physical we canvassed earlier.

If the two questions 'what is the relationship between mind and body?' and 'what is the relationship between the mental and the physical?' are different, it becomes a matter of some interest to see why they have often been treated as more or less the same. One answer which seems plausible is that the intuitions people have about the possibility of disembodiment stem from intuitions they have about the nature of the mental and the physical. Whatever one's feelings about the advances of biology and medical science, it is still pretty clear that we are ignorant about many of the properties of our bodies. Moreover, in days gone by we were even more ignorant than we are now. In this state of bliss, one would have thought that people would be very hesitant about claiming that they, that is, their mind, might be disembodied, yet still continue to exist. Normally, one would expect that the grounds for such a claim would be that someone had a list of all the constituents of the body, and all the constituents of the mind, and recognised that the second list contained no items on the first list. But we are surely not in such a fortunate position as to possess both lists. So, why are so many convinced that

their mind may exist separately from their bodies?

The answer cannot be in the different access we have to our minds and to our bodies, because to put the matter this way is to presume what we are not entitled to presume. It is not to be doubted that there is some sort of phenomena that goes under the name of introspection. What is crucial to recognise is that for all that has been said so far, introspection may yet be another way of discovering bodily properties not hitherto noticed by other means. Therefore, it must be the peculiarity of mental properties, in the eyes of many, which convinces them that their minds may continue to exist disembodied. These people must think that the character of mental properties somehow makes plausible the thought that the constituents of the body could not possess such properties, or some thesis of this sort.

The explanation of the conflation of the two questions we have been discussing brings us to the first constraint upon any account of the physical. It is that we should try to provide an account which makes sense of the traditional debate centred around the mind-body problem. It must at least be conceivable why people thought, and still do think, that the mental is very different from the physical, given our demarcation of the physical. For it to be conceivable, the account we give of the physical should be such as to make it possible for

people to believe that the mental does not fall within the category, whether or not this is true, or make it possible for people to hold that the physical is to be demarcated in a way which does establish that the mental is not physical although this belief as to the character of the physical may be mistaken as far as our account is concerned. Putting the point briefly, any reasonable account of the physical should make sense of our possible misconceptions of the mental and physical. By doing this, we will explain, to some extent, the intuition that some people have that they may exist disembodied.

It may be asked whether an adequate account of the physical must make it reasonable for people to believe that mental things are not physical things, or for people to believe that the concept of the mental is not identical with the concept of the physical.³ Surely, the answer is that an adequate account must first make it reasonable for people to believe that mental things are not physical things. The traditional debate is about the nature of the world and not the nature of our concepts. It is not to be denied that a more modern debate has grown up around the concepts of the mental and the physical, but this seems largely to be a consequence of philosophical methodology. By becoming clear about the nature of the concepts, we will become clear about the relationship between mind and body, it is sometimes said. This, if true, is satisfying

to a philosopher because conceptual inquiry is so much a part of his or her stock in trade. Nevertheless, this is a subsidiary issue.

We may recognise this point in another way. If one assumes that our concepts of the mental and the physical are derived from our experience of the relevant entities, then the explanation of the character of the concepts will derive from the nature of the entities that we experience. So, although the distinction between concepts and things, utilised in the question we are considering, is legitimate, we may, for the reason just mentioned, hold that an explanation of the debate about the nature of the things, will in turn be an explanation about the nature of the concepts. From our point of view, it is the former which is of importance. We now turn to the second constraint.

Sometimes it is assumed that the questions 'what is the relationship between the mental and the physical?' and 'what is the relationship between the physical and the non-physical?' are much the same. But, this would only be true if one assumes that all mental things are non-physical and that the mental exhausts the non-physical, so to speak. However, to make such assumptions at this point is both unwarranted and undesirable.

In recognising that the assumption just mentioned should not be made, we come to the second constraint upon

any account of the physical. The argument at the beginning of the chapter contained, as one of its premises, the claim that mental events are not physical. We need to develop an account of the physical which allows us to assess such a claim. For the assessment to be unprejudiced we should not develop our understanding of the physical by, first, identifying those entities which we definitely take not to be physical, and then developing an account that arrives at this conclusion. Mental events are going to be candidates for what many take to be not physical. To take them as such to begin with would hardly establish the credentials of our inquiry as unbiased. What we should do is identify these entities which definitely are physical and try to find a feature or features of them which, on the one hand, minimise our commitment to holding certain things not to be physical, and, on the other, shows the interest and viability of such a categorisation. When we have developed an account of this sort we should then and only then turn our attention to the mental.

To summarise, two constraints have been argued to govern our acceptance of any attempt to characterise the physical. The first constraint was that:

any account of the physical should make intelligible the misconceptions or otherwise, which have lead to the traditional dispute about the relationship between mind and body.

The second constraint was that:

any account of the physical we offer should not prejudice the issue of whether the mental is physical, while, at the same time, isolating those features of what we definitely take to be physical and displaying their interest.

It is to these that we shall appeal in the sections that follow.

(c) Various characterisations of the physical that have been previously offered.

(i) Definition in terms of Physics

A number of philosophers⁴ recently have supposed that we should understand physical entities to be first and foremost those identified by correct physical theories, a physical theory being defined as any theory which is offered as part of the scientific enterprise of physics. Clearly, we cannot count as physical only those things which are identified by correct physical theories; that is why we said that physical entities were first and foremost those identified by correct physical theories. There will be some things which we intuitively count as physical, chairs for instance, which will not be identified by physical theory. Also, physics is not the only science which speaks of physical things, both chemistry and geology do among others. So, presumably the

idea must be that if something is physical, it must in some way be related to the items identified by physics.

The phrase 'first and foremost' in the context in which we used it may suggest that we are talking of entities which have some sort of special claims to existence or whatever. Some people⁵ do seem to view physics as fundamental in this way. Nevertheless, the phrase 'first and foremost' need not and should not be interpreted in that fashion. All it is meant to indicate is that the entities identified by successful physical theories have some central role in the definition of the physical.

How are we to relate the entities identified by correct physical theory to those entities which other sciences talk about, and to non-scientific entities such as chairs? If we allow our attention to sway from the last mentioned type of entity, we might be inclined to claim that if a science could be reduced to physics via what are commonly called 'bridge laws' then the science would concern physical entities. It would then be, according to this view, legitimate to talk of these sciences as 'the physical sciences'. However, various problems beset the notion of reducibility and we would still have to deal with non-scientific objects such as chairs. The following account seems more suitable.

An entity is physical if and only if it is either

mentioned by correct physics in one of its theories,
or composed from entities so mentioned.

It will be this version of the idea that the physical may
be defined in terms of physics which we shall consider.

First, let us savour its good points. One advantage
of this account is that it is suitably empirical. We are
not, as philosophers, proclaiming what the physical is
from our armchairs, but rather letting the world gently
inform us through experimental results and the theories
scientists develop to account for them. A second
advantage is that it satisfies the two methodological
constraints we introduced in the previous section. We
have not, in line with the second constraint, committed
ourselves to counting certain things as definitely not
physical. It is by no means clear, for instance, that
such a definition rules out the mental being physical.
Equally, the definition renders understandable the debate
about whether the mental is physical, in line with the
first constraint. Those who assert that the mental cannot
possibly be physical, because the physical is chunky and
the mental is not, are, if they are in error, in error
because they have an understandable but inadequate
conception of the physical. They have a certain view
about the type of entities which will be mentioned in
correct physical theories that is false. Nevertheless,
even though there are these two advantages to the present

account of the physical there is good reason to think it ultimately inadequate.

The first problem with it stems from our stipulation that what was physical was identified by, or composed from that which was identified by, correct physics. Suppose in correct physics, there are no particles mentioned. Then, it would follow that our conception of a particle is a conception of something not physical. Yet, surely this is wrong. Even if some particles are conceived to be very strange, there are many which are rightly conceived as paradigmatically physical. To rule these out as physical things just because they do not occur in the world seems unacceptably severe. Our notion of what is physical is broader than that.

To alter the demarcation of the physical by dropping the qualification 'correct' and talking instead of what is currently identified by physics, does not resolve the difficulties an account of this kind faces. For then we face a problem we have avoided so far. Suppose that we are wrong about the fundamental constituents in the universe, there are no particles. Then, every object which is composed from these other fundamental constituents would not be physical. A chair, for instance, would not then count as physical.

A further alteration is obviously needed. Let us read the account as utilising the notion of any physics.

A future physics, a past physics, a present physics, a correct physics, all of these demarcate entities which are physical. But this still does not classify all the physical things. Nobody has proposed, we may hope, that the basic constituents of the universe are very small transparent pyramids. These pyramids, it must be emphasised, are not the composite entities we come across with that shape, they are fundamental pyramids. Must we suppose that since no physicist will propose such an entity in a theory, these pyramids are not physical? Once more the conclusion seems unintuitive.

There is a final objection. It is reasonable to suppose that each science, has what we might call proprietary causes, that is entities of a type for which there is a commitment by the science to propose them as causally explanatory of a certain range of phenomena which the science aims to explain. Thus, we might expect for physics that there is some generally accepted notion of the sort of entity that it is reasonable to postulate as explanatory of the phenomena with which the science is concerned. Talk of particles and waves, however extended this talk is from our every day understanding of these notions, fortifies such a claim. If we suppose that the only account of the physical is the one we are presently considering, the supposition is tantamount to allowing that nothing constrains physicists in the type of

entities they postulate, to explain the phenomena with which they are concerned. Instead, they have only a distinctive range of phenomena which they seek to explain, and community practice upon which to draw. In this science, unlike others, the community practice does not lay down the type of entities which may be postulated *per se*, but instead teaches the physicist how to think of a problem. Then the physicist excogitates the sort of entity which will be appropriately explanatory, yet this excogitation is not sufficiently principled so as to allow us to talk of the sort of entities which are generally counted as physical.

The inability to provide a principle describing what type of entities physicists will postulate, and thereby what type of entities are physical, opens up this version of the first account of the physical to an objection stemming from our first constraint. The claims of those who say the mental is distinct from the physical and those who say that it is not must be counted as predictions about the behaviour of future physicists. However, not only does the content of, and arguments offered for, these claims bear no witness to the fact they are predictions concerning the behaviour of physicists, but also, in the light of this, it would be hard to see how the debate has been a rational one.

Of course such a view of physics cannot be argued

against with any ease. A careful scrutiny of the science and the workings of its disciples would have to take place. Nevertheless, it seems implausible, especially in the light of previous objections. Moreover, if we could give some principled account of the nature of the physical, this would be counterevidence to the claim that physicists did not systematically postulate a certain type of entity, on the supposition that the account is at least right in one respect, that being that physicists do postulate physical entities. At any rate, it is to the principled alternatives that we turn. Some time has been spent discussing this proposal, because it is so often mentioned by philosophers pressed for an account of the physical. There are, however, other candidates.

(ii) The physical as spatial

A long tradition has emphasised that the physical should be understood as in some way spatial. Descartes was a famous exponent.⁶ There are at least two ways in which we may try to develop the idea of the physical as spatial. The first emphasises that for something to be physical, it should be located in space, the second requires in addition to location in space, occupation of space. It is unfortunate that we cannot go into the notion of space here, our understanding of it will have to be intuitive. Instead we will examine whether either

proposal is a plausible characterisation of the physical, whatever space is.

The problem with saying that anything which is physical is located in space is that the traditional debate between those who take the mental as a distinct category from the physical, and those who do not, becomes hard to understand. While it may be true that if one is in the grip of the conviction that the mental is peculiar, it then becomes increasingly difficult to envisage it located in space, it certainly does not seem that the intuitions that people have that the mental is not physical stem from a conviction that the mental is not so located. A simple observation seems to establish this claim. It is that *prima facie* it is easy to locate the mental. The mental, or more specifically, minds, are where people are and a person's body usually provides a good indication of that person's location. Nor can the worry that minds may not be precisely located cut much ice. Apparently, some particles accredited to be physical are difficult to precisely locate. Of course it is not our aim to demarcate the physical so that the mental is definitely not part of it. What is crucial, and it is to this point that the remarks just made are addressed, is that the demarcation render plausible the controversy.

On this front, the notion of the physical as space occupants seems much more creditable. The difficulty of

appreciating what it would be for a mind to occupy space at least makes the intuition that the mental really is different from the physical easy to comprehend. The problem with this proposal is that it would count as non-physical particles identified as point masses. These particles have some mass but occupy no space. I am not quite sure how acceptable the theories which postulate such entities are, from a scientific point of view. But this does not really matter. What we should consider is whether if a theory did propose that there existed such things we should conclude that there existed non-physical things.

Obviously, point masses are not something one can visualise. If it were a necessary condition of the physical that it could be visualised, this would suggest that they were not physical. However, we have had to grow used to the idea that what physics describes, we cannot visualise. In addition, point masses have a property which we would intuitively count as physical, mass. Yet, its failure to occupy space when instantiated would categorise it as non-physical. So, unless we are stuck with no other account of the physical than this one, it seems reasonable to view the present account as unsatisfactory.

Each of the accounts of the physical in terms of space we have considered look inadequate. But, some part

of the intuition that the physical can be linked to the spatial may be saved. It certainly seems that a necessary condition of something being physical is that it is located in space. Equally, if something occupies space, one is inclined to take this as sufficient for it to be physical.

(iii) The physical as spatial and causal

One could be forgiven for thinking that an account of the physical just in terms of space, fails to distinguish circumscribed regions of empty space, or points in space, from the entities which we are trying to demarcate. The claim that space occupancy, for instance, is a sufficient condition of the physical, has been thought to be only true because a tacit interpretation of occupancy was subscribed to which did not take merely the geometrical specification of a region of space as a space occupant. The natural question to ask is what extra is there to the physical? What fills space or is located in such a way that it is physical? One answer, which has been given, is that the additional component should be understood in terms of causality. Thus we get:

An entity is physical if and only if, first, it has spatio-temporal location and, second, it has some observable property or causal power.

The disjunction 'observable property or causal power'

needs some discussion.

Some who put forward this account, Rudolf Carnap and Ted Honderich,⁷ constrain the application of 'causal power' in the following way. They hold that the notion of an observable property, a property accessible to our senses, is central to differentiating between empty space and what is physical. They then loosen this constraint, to allow in 'non-observable' items, but only to the extent that if these items have a causal relationship with what is observable, they too will count as physical. They do not allow that something is physical if it has a causal power but somehow does not effect our senses. This seems unsatisfactory. It is true that we will never apply the predicate '- is physical' to an entity which does not in some way make itself known to us, but this does not rule out the applicability of the predicate to entities which fail to make themselves known so long as they share the relevant characteristics with those which are known and are physical. The obvious defence to this criticism is to say that part of our notion of the physical is of something which interacts with the senses, even if at one step remove, but this does not really seem to be part of the notion.

A second objection to the above account will be considered shortly. First, let us amend the Carnap-Honderich version of the account we are examining.

Something now may count as physical if in addition to spatio-temporal location, it has a causal power full stop. It does not matter whether this causal power is a power to affect the senses or not. On one interpretation of Anthony Quinton's⁸ writings upon this subject, he adopts such an account. The second objection, now, is this: upon the account we have before us, anything which is causally inert is not physical. Should one be prepared to adopt such a commitment? It seems reasonable to say that there are arguments on either side. No doubt it is difficult to make sense of a causally inert thing. In particular, one cannot imagine it. But then one has often been told that one should not try to imagine the entities which physics talks about yet they are generally supposed to be physical.

It might be argued that there would never be any reason to postulate such an entity. But this is not entirely true. Suppose there were, on the part of scientists, a commitment to some relevant conservation principle, let us dub it 'the conservation of the physical'. If as a result of a change in the universe, it looked as if the total amount of physical things had decreased, the principle could always be preserved by the postulation of causally inert entities. There might be good scientific reasons in other areas for wanting to maintain the principle, so such an approach would not be

unacceptably ad hoc.

Even if the account of the physical is unaffected by the previous objection, it does not perform well with regard to our methodological constraints. The account can make no sense of the traditional dispute between those who suppose the mind is non-physical, and those who deny it. What mistake could these disputants possibly have made, and still be making? It is relatively easy to show that the mental is physical as far as the present account is concerned. All one has to do is note the following two things.

(a) Some mental events cause behaviour

(b) Behaviour is observable

Those mental events which do cause behaviour are therefore physical. Anyone who wants to preserve the non-physical nature of some mental entities would have to adopt some form of epiphenomenalism. They would have to claim that there are some mental entities which do nothing except possibly affect other mental entities which do nothing as far as the subject's behaviour is concerned. Yet, a prior commitment to epiphenomenalism cannot be the reason why people have held the mental is not physical.

In fact, the account of the physical we have latterly been considering is subject to an even more devastating criticism, if the methodological constraints we have adopted are acceptable. One does not even have to point

to the effects mental events have on behaviour to establish that they are physical. So long as they have causal influence upon each other, it will follow that mental events are physical. Even the most ardent friend of the proposal that the mental is physical should not be happy with so easy a victory.

(iv) An epistemological criterion of the physical

There has been a tendency to see our epistemological access to the mental as distinctive of it.⁹ This might seem to allow us to provide an account of the physical which contrasts with what is distinctive of the mental. If the contrast was such as to rule out the mental being physical, such an account would obviously be unacceptable. However, there is no reason why the respective accounts of mental and physical should be drawn up in such a clumsy way.

It will not be necessary for us to go into great detail as to the character of the accounts of the mental and physical based upon epistemological access for us to see why they seem unsatisfactory. We shall understand the notion of a distinctive epistemological access to the mental as meaning one of the following: either, that we come to beliefs about our own mental lives without evidence, or that we have superior claims to be correct or be justified in our beliefs with regard to our own mental

lives, or that we have a special means of access to our own mental lives and not others which some call 'direct access'. Some combination of these may be true. At any rate, we may then define the mental in the following way.

An entity is mental if and only if one and only one person has a distinctive epistemological access to its existence and/or nature.

If one then defined the physical as follows:

An entity is physical if and only if it is not the case that one and only one person has a distinctive epistemological access to its existence and/or nature.

We would infringe the second methodological constraint that we adopted earlier. By definition, the physical would be identified as that which is not mental.

An alternative, and preferable, way of formulating the criterion of the physical would be to put it thus:

An entity is physical if and only if it is of a kind such that it is not the case that for each member of the kind there is unique to it a person who has a distinctive epistemological access to its existence and/or nature.

Defining the physical in this way leaves it open whether mental entities, according to the criterion we have offered, are physical. A mental entity would count as physical just if it belonged to a more inclusive category

that included entities whose existence or nature were not the objects of the epistemic privileges identified earlier. Of course, some might argue that no sense can be made of the notion of epistemologically privileged access without either supposing that certain intuitively physical events are non-physical, or that certain mental events are definitely physical, so infringing the constraints that we have identified earlier on. We shall not offer such objections however. It may well be possible that a suitably sharpened notion of such access can steer between these difficulties.

The objection to the present account that we shall offer is that it gives us little insight into the nature of the physical by the demarcation offered of it. Surely, we want to know what it is about the physical that results in some members of the category not being so related to one of us that that person can have a distinctive epistemological access to it. We would want to know, for instance, what feature mental entities would have to possess to belong to the more inclusive category identified in our criteria of the physical. Yet, we are left in the air over this matter.

(d) **The Positive Account**

We saw that there does seem to be something to the idea that it is the absence of a special type of access to

the physical, in general, which is in part characteristic of it. In the light of this thought, and the preceding remarks which have been made about other accounts, the following demarcation of the physical is intuitive.

An entity is physical if and only if,

(a) It is spatio-temporally located, or composed from entities that are;

and (b) Its existence does not imply that a subject is aware of it.

The reason for adding that something may be physical if it is composed from entities that are is to allow that the universe may be physical even if it does not make sense to suppose that it is located in space and time.¹⁰ Obviously, it is the second condition of this account that is going to be controversial, if only because it may seem rather opaque. It is this, therefore, that we must discuss.

The notion of awareness adverted to by the phrase 'awareness of' should be taken to create an 'extensional context'. In the descriptions of the contents of states of awareness, co-referring or co-extensive terms can be substituted for each other without altering the truth value of the descriptions, assuming them to be sentences. For example, I am aware of your favourite tree in the relevant sense even if I am not aware of it as your favourite tree. Applying this point to the present case, we may say something is not correctly categorised as

physical just because we are not aware of it as the thing it is, although we are aware of it and could not but be aware of it. Apart from this, there is no particular type of awareness which is being appealed to in the present account.

What does it mean to say that there is something whose existence implies that a subject is aware of it? In the case of propositions, we say that one proposition implies another if it could not be the case that the first proposition is true and the second not true. Turning from propositions to the matter at hand, the idea seems to be that:

An entity A's existence implies a subject's awareness of it if and only if it could not be the case that A exists, and a subject not be aware of it.

Now, it may be thought that this notion is incoherent. Propositions can imply propositions, but the natures of things in the world cannot imply other things. One can but agree. However, the basic idea can be formulated in terms of propositions. One may say that 'An entity A's existence implies a subject's awareness of it' is an abbreviation for 'Any proposition that has as a necessary condition of its truth the existence of the entity, implies any proposition that has as the sufficient condition of its truth the awareness of a subject'.

Although we have specified the character of a certain

sort of entity in terms of implication it may be better if we introduce more explicitly metaphysical notions to articulate what we have in mind at this point. The most important of which, for the discussion that follows, is that of 'metaphysical sufficiency'. This notion is the metaphysical equivalent of causal sufficiency. A preliminary analysis of it is as follows.

An entity A is metaphysically sufficient for the existence of an entity B only if, necessarily, if A exists, then B exists.

The use of 'necessarily' here and hereafter should be taken as indicating metaphysical or logical necessity indifferently, in line with the policy of being non-committal that we adopted in the introduction. It must be fairly apparent how metaphysical sufficiency so analysed relates to the notion of implication specified above.

So far, we have only provided a necessary condition upon one entity being metaphysically sufficient for another. The reason why what we have specified as a necessary condition is only a necessary condition is that the notion of metaphysical sufficiency is plagued by the same problem as we shall later see plague that of causal sufficiency. In the present case, it is what we may call the problem of spurious metaphysical sufficiency. Suppose that I am presently aware that you are situated one foot away from me to the left. Then it seems that you possess

the relational property of being one foot away from somebody who is aware of this fact. Now, consider that relational property. It seems inappropriate to say that the relational property is metaphysically sufficient for the awareness. There is no way in which one can truly say that the relational property 'brings about' the awareness of the subject. Yet, if we supposed that the necessary condition supplied was a sufficient condition as well, then that is precisely the sort of thing we should be prepared to say.

One may try to capture this point by offering the following account.

An entity A is metaphysically sufficient for the existence of an entity B if and only if

- (a) Necessarily, if A exists, then B exists
- (b) It is not the case that condition (a) is met because the existence of entity A is metaphysically dependent upon the existence of entity B.

The problem is that it may seem that a claim of metaphysical dependence should be understood as a claim that A, the entity that is metaphysically dependant on B, could not exist unless B exists. The latter thought is generally taken to be equivalent to necessarily, if A exists then B exists. So, it seems that (a) and (b) are

in flat out contradiction to each other. What we must do is accept that the notions of metaphysical dependence and sufficiency cannot be fully articulated and accept that the usual formulations provide only a necessary condition for their occurrence. As it seems reasonable to suspect that the notion of metaphysical sufficiency is more controversial than that of metaphysical dependency, we shall take metaphysical dependency as primitive and define metaphysical sufficiency in terms of it. As we have done, in the analysis offered above. However, given that the notion is controversial, it would be helpful if much of what we say rests upon only the weaker notion captured by condition (a). This is what we shall try to do and is the principle motivation behind talking in terms of 'implication'.

Our account of the physical had as a necessary condition, that such an entity's existence does not imply a subject is aware of it. Why is this thought to be an intuitive condition to place upon the physical? There are two answers to be given at this point. The first concerns Hume's Principle of Distinct Existences.¹¹

Hume is well known to be committed to the view that if two entities are distinct, then they cannot have a metaphysically necessary connection between them. His establishment of this claim, to his satisfaction, seems to be among the grounds for his further conclusion that there

is no causally necessary connection either, although it has become controversial whether he does, in fact, arrive at this conclusion.¹² At any rate, it would not be too much of a distortion of one of Hume's doctrines to suppose he believes in something like the following principle.

Necessarily, if two entities A and B are distinct, then neither necessarily, if A exists then B exists, nor necessarily, if B exists then A exists.

The account of the physical we have offered adheres to this principle. Moreover, it is not hard to see why we should suppose that the physical satisfies it. Two distinct physical entities, one is inclined to think, take up two distinct spatio-temporal locations. Given that they do, it is surely then conceivable that one exist without the other. Taking conceivability to be a good indication of what is possible, it seems to follow that either may exist without the other.

Now, it is true that we have not ruled out non-physical things having spatio-temporal location. However, the thought is that we would be satisfied to call anything physical which possessed spatio-temporal location and satisfied Hume's Principle of Distinct Existences. Any entity which did not satisfy the principle might be considered sufficiently strange to count as non-physical even if it did satisfy the first condition concerning

spatio-temporal location. However, we cannot be sure. What is plausibly more certain is that something would be non-physical if it infringed Hume's principle by implying that a subject is aware of it. The motivation for this more liberal understanding of the physical will become apparent when we return to our discussion of the claim that the physical realm is causally closed.

Clearly, if there are entities that count as non-physical, our remarks concerning why the Principle of Distinct Existences is true would be incomplete. It cannot just be the occupancy of distinct spatio-temporal locations that renders the existence of the entity independent of the existence of the other. However, our remarks at that point were to illustrate the connection between a natural understanding of the physical and the Principle of Distinct Existences. The existence of counterexamples to the principle would indicate that it plays a non-trivial part in our understanding of the physical in that it is only true of distinct physical entities.

The second reason for adopting the requirement that a physical entity's existence does not imply an awareness of it concerns the notion of objectivity. It seems reasonable to suppose that physical entities are objective. For our present purposes, the following analysis of objectivity will suffice.

An entity is objective if it could exist with the character it has regardless of whether any subject (or collection of subjects) is aware of it.

To say something could exist regardless of whether something else is the case is to suggest that it is neither metaphysically dependent upon, nor metaphysically sufficient for, it. Thus, we see that, by denying that a physical entity is metaphysically sufficient for a subject's awareness of it, we thereby imply that, in the sense we specified, physical entities are objective.

This raises the question of whether something which is metaphysically dependent upon, rather than sufficient for, a subject's awareness of it is non-physical. It seems reasonable to suppose that this is sometimes the case, namely, when the entities in question are subjective. However, it is difficult to specify when an entity is subjective because we come face to face once more with problem cases such as that of the relational property of being one foot away from somebody who is aware of this fact. That is why we only provided a sufficient condition of the objective. All we can therefore do is utilise it to provide a necessary condition of the subjective.

It may be possible to develop a notion of subjectivity in terms of the notion of the nature of awareness being metaphysically sufficient to determine the

nature of a 'subjective' entity, but, a discussion of this would take us too far afield. It will not be necessary to pursue these matters further, for the argument that follows. The crucial point is that we take physical things to be a species of the objective. Thus, we can see a second reason for adopting the more restrictive account of the physical, that which says that the existence of physical entities does not imply that a subject is aware of them. It is this account that captures the intuition that should subjective entities exist, they would not be physical. We must just be sure that we do not appeal to the features of it that misclassifies our relational property as non-physical.

A question that was asked with regard to another account of the physical, discussed earlier, was 'What distinguishes empty space from physical items?' The answer that our account gives is that nothing that is essential to the characterisation of the physical does distinguish in this manner. Points in space are counted by our account as physical too, and this is surely intuitive. Instead, particular ways of being objective in space differentiate empty space from other physical items, for instance, possessing mass. In taking this attitude, we clearly do not endorse the line of criticism offered against the claim that the physical is merely the spatial, namely that it fails to differentiate empty space from the

physical.

No doubt the account is not as clear as it should be, but let us move to possible objections to it, rather than attempt to refine it. How it fares with the objections may make its motivation and character clearer.

The first objection is that of extravagance. It may be claimed that the notion of something having its existence imply that a subject is aware of it is sufficiently incoherent to make every entity physical. In reply, it seems reasonable to point at the nature of the epistemological privileges some would claim for the mental. For some, it makes no sense to envisage certain mental entities as ensconced in a subject's mind, without that subject being aware of them. The felt quality of pain is an oft cited example. If one takes this felt quality to be physical it would seem possible to envisage that quality occurring independently of any awareness of it. The putative impossibility of envisaging this situation may be based on the thought that the quality is one whose existence implies an awareness of it and this same thought may plausibly be taken to be at the root of some people's objection to the claim that everything is physical. Whether or not they are right in this objection does not concern us here. What does is the existence of a line of thought which finds a notion to which we are appealing coherent, moreover, a line of thought which is

concerned with the nature of the mental.

Our second methodological constraint upon any account of the physical was that we should not prejudge the issue as to whether the mental is physical or non-physical. Consequently, we are committed to the view that it is not obvious that mental entities have an existence that implies that a subject is aware of them. This gives rise to a second objection,¹³ that we offend against the constraint. It is obvious that the mental is not physical on this account.

In reply, we can say that it is just not obvious that the mental is non-physical according to the account we have adopted. Some philosophers would quite properly consider it an open question. David Armstrong¹⁴ has explicitly stated that mental entities can exist yet a subject be unaware of them. So it seems we are therefore entitled to suppose we have not prejudged the issue in the eyes of some. For the more doubtful, all we may do is appeal to the intuitiveness of our account of the physical. Do we really want to say that something which fails the second condition is physical?

The third objection concerns those physicists and philosophers who argue that Quantum Mechanics, among other things, establishes that something of the order of Idealism, Phenomenalism, or Kant's Transcendental Idealism is true of the world, each of which having the upshot that

what we call the physical in some way implies an awareness of it. Surely, the objection goes, if these physicists are right, then the physical would be shown to be other than our demarcation of the physical would suggest. Thus, the demarcation is wrong. In reply, one might say that if physics did come to the conclusion that one of these philosophical theories of the nature of the world was right, then physics would have come to the conclusion that there is nothing physical. Our conception of the physical ought to be separated from the theories of physicists. If the line of reply is thought acceptable, then there is yet another objection to be had with regard to the account of the physical in terms of physics. The separation of our conception of the physical from the subject matter of physics, may not, for obvious reasons, result in physicists freely postulating mental entities to populate the universe if these mental entities are really not physical. But, there is no clear case for saying that the reason why they should not, is that physics is concerned with the physical. Much more important a reason for neglecting to postulate mental entities, in the case envisaged, is that their character is so opaque that they are of dubious theoretical use.

The fourth objection arises from the case of dispositional and unconscious mental states. They exist when we are not aware of them. So it would seem to follow

that they must be physical. Does this not make little sense of the traditional debate? It would be ironic if a criticism with which we continually prosecuted others, came home to roost in our own account. But, fortunately, this does not seem to be the case. For one thing the traditional debate has never been centred on unconscious and dispositional mental states. They have always been an embarrassment. So, it would be quite possible for us to make sense of the traditional debate in the limited terms in which it was usually conducted, and admit we had some reason for thinking that unconscious and dispositional mental states were physical. There are other ways in which we may defend the account. First, it might be claimed that something could only be an unconscious or dispositional mental state if necessarily it either gives rise to conscious mental states at some point or has the capacity to give rise to conscious mental states. In which case, although there was some reason, based on our definition of the physical, for thinking that unconscious and dispositional mental events are physical, the matter is complex. There is some metaphysical or logical connection between the existence of these mental events and awareness which may easily have been mistaken for that which is required to make these mental events not physical. Second, it may be thought that our account of the physical should be tightened to rule out entities

whose existence is connected in any way with our awareness of them, and the resultant controversy over this matter could explain the traditional dispute. Third, we might hold that unconscious or dispositional mental events could not exist, if awareness of some mental events did not exist, and hence that their existence implies some awareness although not of themselves. Whatever line is taken, since the traditional debate can still be made sense of even in this area, the objection to our account looks less than compelling. It does indicate that the account needs more work, but, for our purposes, it will not be necessary to come down on one approach rather than another.

This ends our discussion of the objections to the positive account. It is to be hoped that the demarcation it offers makes some sense of the traditional dispute introduced at the beginning and does not prejudge the issue of whether the mental is physical or non-physical, so satisfying the methodological constraints we adopted. It is also to be hoped that the question 'What is the relationship between the mental and the physical' is seen

to be well formed and substantive. We shall use this account to try to assess whether mental events are physical, and thereby to assess whether their explanatory role in the explanation of behaviour is threatened if we hold that the physical world is causally closed. However, before we can do this, we must refine our understanding of the argument formulated at the beginning of the chapter. This we shall do in the next chapter.

References

1. To name a famous example, D. Davidson (1970), 'Mental Events' (in his, Essays on Actions and Events).
2. David Hume (1739/40), A Treatise of Human Nature (I, iii, 2-8); (1748), An Enquiry Concerning Human Understanding (Ss 4,5).
3. It was, by David Ruben.
4. c.f. J.J.C. Smart (1963), 'Materialism' Journal of Philosophy 60; D.M. Armstrong (1977), 'The Causal Theory of Mind' (in his (1980), The Nature of Mind); D. Lewis (1983), 'New Work for a Theory of Universals' Australian Journal of Philosophy 61; H. Putnam seems to work with this notion of the physical, for example, see his Reason, Truth and History (p.78).
5. c.f., on one interpretation, W.V. Quine (1960), Word and Object (Ch1).
6. c.f. R. Descartes (1644), The Principles of Philosophy (Pt.2), (1641), Meditations on First Philosophy (II and VI)
7. R. Carnap (1928), The Logical Structure of the World (ss.18); T. Honderich (1988), A Theory of Determinism (p.87-89).
8. A. Quinton (1973), The Nature of Things (p. 46-53).
9. Among those who offer an epistemological criteria of the mental are: J. Kim (1971), 'Materialism and the Criteria of the Mental' Synthese 22; A. Levison (1983), 'An Epistemic Criterion of the Mental' Canadian Journal of Philosophy 13.

10. I owe this point to Richard Sorabji.
11. See footnote 2, especially Hume, A Treatise of Human Nature, (Part III s6).
12. e.g. G. Strawson (1989), A Secret Connexion; E. Craig (1987), The Mind of God and the Works of Man (Chs 1 & 2).
13. As Ted Honderich pointed out to me.
14. D.M. Armstrong (1968), A Materialist Theory of Mind, and D.M. Armstrong in D.M. Armstrong and N. Malcolm (1984), Consciousness and Causality, (p.124-135, p.205-9).

Chapter 2

Refinement of the first argument

In the first chapter, we sketched an argument that had as its conclusion that mental events do not have a role to play in the psychological explanation of behaviour. Two of the premises involved, in part, an appeal to the notion of causation. So, we should consider whether we may keep our understanding of this notion at an intuitive level, or whether we should try to refine it. Sometimes it is good philosophical practice to develop an argument and claim it is independent of specific interpretations of key phrases. Being deliberately non-committal in certain areas is a good way of persuading people of something while avoiding controversy. Unfortunately in this case we cannot be deliberately non-committal. As we shall see the nature and extent of the challenge of epiphenomenalism depends upon what understanding of the notions of cause, and of causal relationship we have. Therefore, the first section of the present chapter will concern what it is for something to be a cause, and for it to be causally related to

something else. We will then move on to refine the first argument in the light of the understanding developed.

(a) Causality

(i) The Causal Relation

The causal relation is supposed to hold between two distinct things. We shall stipulate that these things are events, but remain sufficiently flexible as to the nature of events to count the merest instantiation of a property or, as we shall put it hereafter, occurrence of a property as an event. A causal relation holds between two events if there is a certain connection between their occurrence. At its crudest, the idea is that one event caused another if it was necessary in the circumstances for the occurrence of the other. One way of putting this is: if e_1 had not occurred then e_2 would not have occurred.¹ 'e₁' and 'e₂' are dummy names for particular events such that if it is not the case that $i = k$, then it is not the case that $e_i = e_k$. The conditional conveys the idea that e_1 was necessary in the circumstances for the occurrence of e_2 because, as many have noted,² when we talk of what would have happened if things were different we presume that the circumstances that we are discussing are like the present circumstances except in regard to the changes mentioned, and any consequences of them.

The interpretation of conditionals of the form 'if e_1

had not occurred, then e_2 would not have occurred' has always been problematic. The conditionals seem to assert a binding connection of some kind between e_1 and e_2 , but apart from that little is known for certain. One may hope that, at least, the conditionals help to illuminate the content of our judgements concerning such a relation. Even if they do not tell us what the 'reality' of the causal relation is, we may suppose they demarcate what we are inclined to say, should we be convinced a causal relation does hold.

Some philosophers³ claim that the causal relation should be taken as a primitive. They view attempts to analyse it in terms of conditionals as doomed to failure. However, it is difficult to see why they should believe this. The only way we understand the binding nature of the causal relations is in terms of the conditionals that we believe to hold. Equally, our judgement that two events are causally related is more or less based upon the thought that certain conditionals are assertable.

The last point is important for our subsequent discussion. We shall presume that the account of causality developed in this chapter provides grounds for the assertion or denial of certain claims concerning what causal relations hold between entities. Sometimes this will result in a disagreement with what is commonly said. Naturally, when this occurs, it may be thought to be a

consequence of some inadequacy in the account offered. However, unless some reason other than intuition is offered, we shall suppose that our account is correct as to what it suggests we should say and that what is commonly said is wrong. Thus, we are committed not to the correctness of a conditional account of causality per se but to the correctness of the view that the only reason we can have for supposing a causal relation to hold is that we suppose certain conditionals to hold. With this in mind, let us turn to the account itself.

We said that 'If e_1 had not occurred, then e_2 would not have occurred' implies that ' e_1 caused e_2 '. However, this claim may be questioned. One problem any account of causation faces is that of 'spurious' causation. Suppose e_1 causes e_2 and e_3 . Moreover, given the laws there were, and the circumstances that held, e_1 could not have failed to cause both e_2 and e_3 . In this situation, would it not be correct to say 'If e_2 had not occurred, e_3 would not have occurred', which would qualify e_2 for being the cause of e_3 ?

We are being invited to compare two situations: one in which e_2 is subtracted, one in which e_1 is subtracted. In the situation in which e_1 is subtracted, it is clear we are committed to the non-occurrence of e_2 and e_3 . On the other hand, concerning the situation in which e_2 is subtracted, we can say one of two things. First, we can

say that e_1 also would not have occurred and hence e_3 would not, but it is because e_1 had not occurred, and not because e_2 had not occurred, that e_3 would not occur. Now we need to cash out the 'because' used in the preceding sentence. An obvious way in which we may do it is by asserting that not all conditionals that relate events in the way indicated imply that these events are causally related. Only those conditionals which are not just a consequence of certain other conditionals count. We would then determine whether or not a conditional was just a consequence of others holding by considering what would happen if we varied the circumstances somewhat. If, in the altered circumstances, the conditional held in the absence of one of the others holding, then the conditional is not just a consequence of the others in the actual circumstances, otherwise, not. Of course, a full account would require us to go into the nature of the alteration, but we shall set such matters aside.

Alternatively, if the approach we have just sketched is unacceptable, we can say that e_1 still would have occurred, and thereby e_3 would have, in the absence of e_2 . Although e_1 brings about e_2 , if we deliberately subtract e_2 , then it does not follow that e_1 would not have occurred. This would involve us in denying that the relevant similar circumstances that we should consider to assess the conditional must keep everything fixed except

the absence of the effect. We shall discuss this approach further under the heading of 'causal priority'.

If e_1 caused e_2 , does it follow that if e_1 had not occurred, e_2 would not have occurred? The answer is no. One reason for this is the matter of overdetermination. One example to illustrate the point is of a driving instructor and his or her pupil. The pupil is about to run somebody over but slams on the brakes. The driving instructor does too, with the duplicate brake peddle. The car stops. As Ted Honderich⁴ points out, as neither action was required for the car to stop the conditional would not hold true for either. Yet, one would want to say that the action of both pupil and instructor caused the car to stop. To incorporate this point we may say that when two events e_1 and e_2 overdetermine another, e_3 , they are still causes of e_3 if the following are true: had e_1 not occurred, then if e_2 had not occurred, then e_3 would not have occurred and, had e_2 not occurred, then if e_1 had not occurred, then e_3 would not have occurred.

The second reason for rejecting the reverse entailment, arises from the recognition that causal relationships are transitive, but the conditionals do not capture this fact. Intuitively, we allow that if e_1 caused e_2 and e_2 caused e_3 then e_1 caused e_3 . Yet, the standard way of understanding the relevant conditionals would not allow such an inference to be made. Now, it is

possible to hold that there are special 'causal' conditionals which do allow inferences of the right kind, but that other counterfactual conditionals do not.⁵ However, this raises problems for those who want to understand the causal relation in terms of conditionals for we do not want the conditionals to be special. Moreover, it is controversial whether even 'causal conditionals' are transitive. An obvious example of this is pre-emptive causation. A chain of events e_1 to e_3 occur such that if e_1 had not occurred, then e_2 would not have occurred, and if e_2 had not occurred, then e_3 would not have occurred. The occurrence of e_1 interferes with the causal path which would have otherwise held between e_4 and e_3 . In these circumstances, we could not say, if e_1 had not occurred, then e_3 would not have occurred, since e_3 would have occurred as a result of the occurrence of e_4 . Nevertheless, it would be true that e_1 caused e_3 .

The situation just described relies upon the same thought as that which comprised the alternative way of resolving the problem of spurious causation, namely, that a prior event e_1 causally related to e_2 would still occur, even if e_2 did not. The most similar possible circumstances in which e_2 does not occur, are ones in which e_1 occurs, but the circumstances are subtly changed so as not to conspire to bring about e_2 . Only if one accepts this, is it possible to hold that if e_2 had not

occurred then e_3 would not have occurred, as required.

A resolution of the present difficulties is to say e_1 caused e_n if there is a chain of events from e_1 to e_n , and for each event in the chain and its neighbour, it is true that if e_k had not occurred, then e_{k+1} would not have occurred.⁶ Unfortunately, one problem remains. Suppose e_1 is the neighbouring event to e_2 in the causal chain and there is some other event e_3 which would have caused e_2 , had e_1 not occurred. Moreover, suppose that e_3 would have been the neighbouring event in a causal chain to e_2 . In those circumstances surely we cannot say what makes e_1 causal rather than e_3 since it is precisely not true that if e_1 had not occurred e_2 would not have occurred.

There seem two options at this point. First, we can allow that a conditional analysis of causation will not be able to capture all the discriminations we need to make with regard to causation but only some. This shall not be the approach we adopt. Second, we can reject the way in which the case is described and say that it is not a genuine instance of pre-emption, but only one of overdetermination. In which case, the only type of pre-emption that we can allow is an event's production of a certain type of event which another would have produced given the chance at another point in time. The remark just made presumes the time of occurrence is an essential feature of a particular event. If not, then pre-emption

of a particular event can occur after all, by its being produced at a different time.

If e_1 is the cause of e_2 it is not the case that e_1 by itself is sufficient to bring about e_2 , other things must happen as well. Consider a particular cause at a certain time. There will be events which must occur at that time, or must have occurred slightly earlier or later, perhaps, for that cause to have brought about the effect it did. Since each of these other events satisfies the same conditions as that we have nominated as a cause satisfies we should be prepared to call them causes too. Admittedly, what we are inclined to call a cause is something especially salient in the circumstances. The lighted match, rather than the presence of oxygen, is said to be the cause of the house fire. It nevertheless seems that this inclination to call one thing rather than another, of the 'collection of causes', a cause concerns how we convey information and contains no fundamental insight into the nature of causation. It is just that we do not report the commonplace.⁷

The complete collection of causes is composed from all events which are individually necessary in the circumstances for the effect at a certain point in the causal chain. They are sufficient for the effect regardless of circumstances if the world is deterministic in this regard, otherwise, and more generally, the

collection of causes are sufficient for the effect to have had the probability it in fact had. Some have rejected the notion of probabilistic causation to which we have just appealed.⁸ But it is better that we do not commit ourselves to any understanding of causality that rules it out unless absolutely necessary.

We are now in a position to summarise our account of the first of the two notions with which we were concerned in this part of the chapter.

A causal relationship holds between e_1 and e_3 if and only if

(a) Either

(i) On one way of resolving spurious causation:

if e_1 had not occurred, e_3 would not have occurred.

On the second way of resolving spurious causation:

if e_1 had not occurred, e_3 would not have occurred, and, the conditional involving e_1 and e_3 is not just a consequence of conditionals holding between e_2 and e_3 and e_2 and e_1 .

(We shall not endorse one or the other here).

or

(ii) if e_1 had not occurred, then if also e_2 had not occurred, e_3 would not have occurred.

or

(iii) e_1 and e_3 are linked by events such that for any two neighbouring events either (i) or (ii) hold.

(b) Both e_1 and e_3 occur.

(ii) Causal Priority

The other notion that we were trying to understand was that of a cause. We have already allowed that there is no significant difference between what we call a cause and other 'causal factors' necessary for the effect to occur in the circumstances. Consequently, the question for us centres on what differentiates causes from effects. We are interested in the question of causal priority, or asymmetry. In particular we will be concerned with whether an account of this notion can be given that makes causal priority some component of the world rather than merely a consequence of how we interact with the world. All anthropomorphic accounts in terms of our experience of human activity and such like will not be considered. That is not because such accounts are clearly wrong. In fact, our discussion will suggest that some such account may well be right. Rather, it is that for our subsequent

discussion the most significant issue will be whether there is a correct 'objective' account.

First, we shall assess whether causes may be distinguished by their location in the account we have provided of causal relationships. In other words, is what we called 'e₁' always the cause? The obvious way to find out is to consider whether condition (a) is satisfied if we take e₁ to be the effect.

Suppose a certain effect had not occurred, what would we be inclined to say about the cause's occurrence? In deterministic cases when we are talking about two events succeeding each other in a causal chain, we would be tempted to say that the cause would not have occurred either, in the most similar possible circumstances. One philosopher, David Lewis,⁹ would have us believe that the cause might still occur, a possibility we first considered with regard to spurious causation. His conclusion depends upon his claim that, intuitively speaking, the most similar possible circumstances would be ones in which the cause is still present but was not sufficient for the effect.

We will not be able to give Lewis' proposal the discussion that it deserves, but a number of points can be made which, at the least, throw into question whether it is ultimately likely to be satisfactory. First, the existence of the intuition that, in some sense, if the

effect were not to occur, the cause would not have occurred, in the deterministic case, is an embarrassment. Unless we are to write off this intuition as a complete mistake, it seems to point to something distinctive about the causal relation. On the other hand, the asymmetry that Lewis finds in the counterfactuals is, first, taken to hold in general. But then it would appear that this latter asymmetry, even if it did successfully differentiate causes from effects, tells us nothing about the nature of causal priority, or whether it is objective. Moreover, there is no reason to suppose that the asymmetry mentioned should be successful.

In his later discussion¹⁰ of the asymmetry, Lewis says it exists because those events we call causes give rise to a whole set of consequences, like a stone dropped into a pond gives rise to many ripples. So, if we supposed that the cause is absent, in addition to one of its consequences, the resulting situation that we envisaged would be radically different from the actual situation and, therefore, not the one which we must consider in our assessment of the relevant counterfactual.

There look to be a number of snags with this suggestion. First, when we envisage the subtraction of a cause from the circumstances, we need not presume that there is no replacement event or events which smooth over the causal gap by giving rise to most of the consequences.

Thus, the circumstances in question need not be radically different after all. Second, even if it is granted that the subtraction of the putative cause will result in many more changes, this alone does not show it is the cause. To establish the latter, it is necessary to show how the fact just cited implies that the event caused the various consequences rather than the reverse. For, we could as well allow that the situation would be radically different because the various circumstances taken together were the cause of the event in question. This may mean that overdetermination takes place more often than we have previously recognised, but perhaps we were just blind to this. What we need to know is why it is correct to look at the matter in the familiar fashion and to this we have been given no answer. It is not enough just to point to an asymmetry and claim that it is the required one.

If we suppose the first conditional of (a) does hold if we take e_1 as the effect and e_3 as the neighbouring cause in the causal chain, it is reasonable to believe that in the cases where the cause is not immediately precedent the third disjunct of condition (a) will hold for e_1 as the effect. So given that our remarks concerning Lewis' proposal are correct, it follows that in the deterministic case our account of the causal relation cannot be used to differentiate between cause and effect.

The same does not hold in the case of probabilistic causation. Once more consider the case of two events occupying neighbouring places in a causal chain. Letting e_1 be the effect makes (a)(i) false. If the effect did not occur, the cause might still have occurred since the cause was only supposed to make the effect probable. A rather satisfactory conclusion for those who have long been recommending that probabilistic causation should be allowed to be legitimate. We have now found that such causation enables us to discriminate between cause and effect in a remarkably clean way. Nevertheless, it does not serve to provide the account that we would wish for since it would have the upshot that the notion of cause would have no application in deterministic cases.

Various other ways have been offered to distinguish the cause from the effect. One way is to take causes as temporally prior to effects. There are four difficulties with this suggestion. First, it has been held by some philosophers, influenced by Einstein's work, that causal priority constitutes temporal priority.¹¹ In which case, we cannot use the latter to provide an independent account of the former. Second, it rules out the possibility of simultaneous causation. Third, it rules out the possibility of backwards causation. Yet, both of these have been argued with some success to be possibilities.¹²

The final objection is a little hard to put. The

intuitive notion of cause has often had associated with it such notions as 'bringing about' and 'making something happen'. These have commendably been considered obscure, and some attempt has been made to illuminate them. But one would hardly want to say that the fact that causes occur temporally prior to effects was the missing illumination. So one must conclude that what is being suggested, by those who speak of it, is a reform of our notion of cause. It might be hoped that a more conservative reform is possible.

An idea put forward originally by Bertrand Russell¹³ and endorsed more recently by Ted Honderich¹⁴ is that causal asymmetry amounts to the following: given an effect of a certain type, and the laws that there currently are, any of a number of types of collections of causes could have existed, whereas given one type of collection of causes, and the laws that there currently are, only one type of effect could have existed. This is an obvious consequence of the way we defined 'collection of causes' earlier, supposing the world to be deterministic. Unfortunately, we differ from Ted Honderich, at least, in allowing that the notion of causality has application in an indeterministic world. So, the suggested asymmetry is not so obviously available to us. There are ways in which we may repair such an account to deal with this problem and these would be worth

considering were the approach not possessed of another, and, from our point of view, singularly unfortunate, failing.

A consequence of the account, that Russell and Honderich favour, is that causal asymmetry is an artifact. Russell explicitly makes this point.¹⁵ It is because of the way we draw up the types of collections of causes, that a particular effect is determined. We could have chosen to specify collections of the events we call effects with the view to showing how they relate to one type of event we have previously called the cause. A number of factors determines our actual choice. No doubt one is that we 'travel' through time in a certain direction. Different ways of travelling through time will result in a different choice as to what are the collections of causes and the effect. What we perceive as causal asymmetry is a matter of temporal perspective. However, whatever truth there may be in this approach it does not satisfy the requirement with which we began our inquiry, namely, that an account of causal priority single out an objective asymmetry.

A fourth account of causal asymmetry takes a distinctively empirical approach to the issue and in this regard resembles Lewis' approach. Some feature of the world is identified as that which, in fact, constitutes the causal connection. Transmission of energy is a

popular example.¹⁶ Then, causal priority is supposed to be determined by the direction of transmission. If I slap your face, the energy in the movement of my hand is transmitted to your face resulting in the displacement of your skin and the slight rocking of your head. The movement of my hand in the appropriate way is supposed to be causally prior to the skin displacement and rocking of your head because the energy is transmitted from the former to the later.

But, are we entitled to come to this conclusion? Let us grant that if energy is transmitted from A to B, then A is causally prior to B. We still need an account of what it is for energy to be transmitted one way rather than another. The problem of direction is just reintroduced in a new context. Our conclusion that the energy transmitted from the hand to the face was based upon a tacit appeal to some other notion of priority, for instance, temporal priority. More generally, it seems clear that any appeal to features of the world that are putatively distinctive of causality and to the asymmetries present in the relations between these features will already presuppose an answer to the problem of causal priority, rather than provide one.

Our brief discussion has therefore ended in failure. We have found no objective means of making out the distinction between cause and effect. Of course, this does

not mean that there is no such means. Nor have we been able to consider all approaches that have been offered, Equally, it may be wrong to seek an analysis of the notion of causal priority. Perhaps it is just one of those notions which should be taken as primitive. However, our discussion does suggest that unless we do take the notion of causal priority as primitive, only a subjective account will be open to us.

This will not matter for most of our inquiry. There, we shall just appeal to our intuitive understanding of the difference between cause and effect, some of which was revealed in the preceding discussion. We shall also allow that two events may be causally related if neither of them are causes. Nevertheless, the conclusions to which we have come, concerning causal priority, will have relevance to what we say in the conclusion.

(b) The argument for epiphenomenalism, premiss (1) again

We are now in a position to come to a better understanding of the first premiss of the argument we discussed in Chapter 1. The first premiss, it will be remembered was this:

- (1) The physical realm is causally closed, in other words, nothing non-physical can have a causal relationship with something physical.

Let us suppose that something should be counted as non-

physical if and only if either it has one or more non-physical properties or it is the occurrence of a non-physical property or a non-physical property itself. Thus, the premiss should be understood as asserting that no entity which has a non-physical property, is a non-physical property or is the occurrence of a non-physical property may be causally related to something physical.

As such, the formulation of the principle that the physical realm is causally closed seems too strong. We should not rule out an entity with non-physical properties being causally related to something physical. What is required is that the non-physical entities do not have influence upon physical entities as a consequence of the former's non-physical properties.

As a consequence, we should alter the rather crude elucidation of the claim that the physical realm is causally closed which follows the phrase 'in other words' in the first premiss. The suggested replacement is this:

- (1)' The physical realm is causally closed, in other words, no causal relationship should hold between two entities, one of which is physical, as a consequence of the other's possession of some non-physical property.

The altered premiss does seem to be what the argument concerning causally closed systems, was aimed to establish. However, to be certain of this we need to

explain what it is for a causal relationship to hold between two entities as a consequence of only some of their properties.

The most immediate answer is to appeal to the notion of a law of nature. Upon this view:

A causal relationship holds between two entities as a consequence of the entities possessing or being composed from the occurrence of particular properties if and only if there is a law of nature which connects these two properties in such a way that they co-occur.¹⁷

Notice that in giving such an account we are not assuming that all causal relationships hold as a result of the properties possessed by the relevant entities being connected by law.

The problem with the suggestion is that it is committed to the claim that a one-off causal relationship between entities could not be taken to be, even in part, a consequence of their possession of properties. But, it is not clear that such a case should be ruled out. On the other hand, if it is said that the account given does not rule out the case because the account deals with causal relationships holding solely as a result of the properties and we are now talking in terms of a causal relationship being 'in part' a consequence of the possession of properties, then the reformulation of the first premiss is

not strong enough. It allows non-physical properties to have some sort of influence.

The obvious solution is to apply our account of causal relations, which made no appeal to laws, straight to the problem at hand. Thus, we get:¹⁸

A causal relationship holds between two entities as a consequence of the entities possessing or being composed from the occurrence of particular properties if and only if the causal relationship between these entities only holds because a causal relationship holds between the occurrences of properties possessed by, or composing, the entities.

In the formulation, 'because' means roughly 'as a logical consequence of'. Some qualifications to this formulation will be made, but, for now, it will serve our purpose.¹⁹

We shall now try to complete our argument for the first premiss understood in the manner we have just articulated. We saw, in Chapter 1, that the presumption that a system is causally closed simplified our search for laws and constrained us in the types of explanations of phenomena we may offer. But, we now need to determine why one should take the physical realm as causally closed rather than any other. Of course, the answer could just be that we believe as a matter of fact that it is. The universe, so we suppose, is entirely composed from physical entities. However, this cannot be the right

answer. Many view that there are problems with claiming there are only physical entities since some are not so obviously physical, yet they hope that these problems can be overcome and the entities in question shown to be physical after all. Why go to this bother? What are the merits in the truth of the claim that everything is physical?

A 'methodological' argument for supposing the physical world to be causally closed is this. There are two constraints we should accept in proposing explanations of phenomena. The first is that the entities we refer to should be publicly accessible, and not just accessible to one person. The second is that the entities proposed to explain the phenomena should be available to explain those phenomena whether or not they are observed. On both grounds it seems arguable whether wholly non-physical entities fit the bill.

If the existence of certain entities implies that a particular subject is aware of them, then it is not clear whether other subjects also have access to these entities. For, whatever is publicly accessible, one would have thought, may exist without implying that any subject is aware of it. We can envisage that that thing to which we have public access may exist independent of the awareness of any one of us, and, if this is so, one may wonder what stops that thing from existing in the absence of the

awareness of everyone of us. So, that of which we are publicly aware seems to be that which is physical.

Equally, if these entities are proposed to explain the phenomena, then some subject must always be aware of the entities in question. From which it follows that if the phenomena occur and no such subject is available, the phenomena in those circumstances cannot be explained in the same way. Yet, can we really allow this to be the case? It is natural to suppose that the absence of a subject to be aware of the explanatory entities does not change the situation so much that a different explanation is required.

The 'empirical' argument that may be offered for the causal closure of the physical world is stronger in some ways, although like the methodological argument, it is suggestive rather than demonstrative. Here, the claim is that we have found good reason to believe up until now that only physical entities are required for explanation. For, suppose something non-physical was needed to explain the occurrence of something physical. We would then expect to find a gap in explanations of our current scientific theories which utilise only physical entities. For example, the activity of the physical constituents of our behaviour would be to some extent inexplicable if something non-physical was to be allowed to be among the causes of behaviour. Yet, many would view the situation

described as scarcely plausible. In particular, because if non-physical entities do have an effect upon physical things, one would expect that this would involve the infringement of certain physical laws and conservation principles which they take to have some degree of scientific respectability.

Those who view this sort of possibility as incredible cite the success that science has had so far in explaining phenomena by postulating physical entities, and take this to be a reason to expect equal success in the future. Moreover, they hold their optimism on this front to be defensible not just because of past successes, but because of present lack of success in coming to any precise understanding of what we would count as typically non-physical entities. They do not consider the alternative a real alternative.

Now, we must agree that it seems overwhelmingly tempting to suppose that physical entities will be all that is needed to explain physical phenomena for the reasons we have just described. Consequently, we have some reason for believing that we should assume, as a methodological principle, that the physical realm is causally closed. It is understandable, therefore, that the opposition to the argument for some form of epiphenomenalism which we are considering arises elsewhere.

(c) The fourth premiss, and the type of epiphenomenalism it supplies.

The fourth premiss of the argument offered in Chapter 1, Section (a) was this:

- (4) If mental events are to have a role in the psychological explanation of behaviour, then they are among the causes of behaviour.

For example, if we want to explain why somebody withdrew their hand from the hot iron rapidly, and with a certain familiar expression, we want to be able to cite something which, as one might say, caused the behaviour described to occur. The explanation we might be inclined to offer is that the person was in pain and believed that the pain he or she felt occurred because of the contact he or she had with the hot iron. It seems reasonable to suggest talk of pain and belief is talk of mental events which were necessary in the circumstances for the behaviour to occur. Had the two mental events not been present, the subject would not have withdrawn his or her hand and grimaced. The collection of causes that would be sufficient for the behaviour would no doubt include other things than the two mental events just mentioned.

If one takes the pattern of explanation provided as indicative of the nature of all psychological explanation one will be convinced that the following thesis is true.

A psychological explanation of those events whose

explanation is in some way proprietary to psychology, will be by the citation of a cause or collection of causes of these events.

It is no doubt from this view of psychological explanation, that the fourth premiss is defensible. Therefore, the argument will go through if the nature of psychological explanation is correctly specified by this proposition.

Suppose for a moment that mental events stand in no causal relationship with behaviour. Then, it would follow from the account of psychological explanation just provided that they play no role in the psychological explanation of such behaviour as the conclusion of the argument offered in Chapter 1 section (a) recorded. Upon one view of what epiphenomenalism means, it would follow from that conclusion that they were epiphenomenal with regard to behaviour. Epiphenomenalism on the view mentioned may be understood as follows:

An entity is epiphenomenal if and only if it stands in no causal relations to other entities.

Now, it might be thought that this definition of epiphenomenalism is too weak. It allows that something which is the effect of something else is not causally epiphenomenal, just by being an effect. This is a matter that we shall return to in the conclusion, suffice it to say that as we could not provide an objective demarcation

of causal priority, it would not befit us to adopt uncritically a stronger formulation. Anyway, nothing that follows depends upon this result. The definition just provided is, also, of blanket epiphenomenalism. As we previously implied, by speaking of 'epiphenomenal with regard to behaviour' more limited varieties of epiphenomenalism may be put forward.

There is, however, another view of epiphenomenalism for which it is an open question whether mental events are epiphenomenal with regard to behaviour, if they stand in no causal relationship to behaviour. This view may be understood as follows:

An entity is epiphenomenal if and only if reference to its existence provides no explanation of the existence and nature of other entities.

The two forms of epiphenomenalism only coincide if you take all explanation to be causal explanation. To demarcate them, we shall call the first 'causal epiphenomenalism' and the second 'explanatory epiphenomenalism'. If the latter is a distinct breed of epiphenomenalism, then of course it follows that even if we find mental events causally epiphenomenal, they may not be explanatorily epiphenomenal.

In the course of this chapter, we attempted to provide a more precise formulation of the argument we first specified in Chapter 1, and offer a defence of some of its premises. The result has been that we have found some reason to believe that if mental events are wholly non-physical, then they are causally epiphenomenal. However, a certain persuasive line of thought suggests that the challenge of epiphenomenalism is far more extensive. It is said that even if we find that mental properties are physical, we shall still find that they are causally epiphenomenal. If this were true, then our focus upon the physical or non-physical nature of these properties will seem misguided. In the next chapter, we shall try to find out whether we are misguided.

References

1. See D. Lewis (1973), 'Causation' Journal of Philosophy 70. The discussion that follows draws heavily on Lewis' account.
2. e.g. David Lewis (1973), reference of footnote 1, (1979), 'Counterfactual Dependence and Time's Arrow' Nous 13.
3. e.g. J. Bennett (1988), Events and their Names (p.50); R. Taylor (1966), Action and Purpose (Chs. 2-3).
4. T. Honderich (1988), A Theory of Determinism (p.18-19).
5. T. Honderich (1988), A Theory of Determinism (Ch.1)
6. This the solution put forward in D. Lewis (1973), 'Causation' Journal of Philosophy 70.
7. J.L. Mackie (1974/80), The Cement of the Universe (Ch.5).
8. T. Honderich (1988): A Theory of Determinism (p.62)
9. See footnote 1
10. See footnote 2

11. See B. van Fraassen (1970), An Introduction to the Philosophy of Time and Space(Ch6) and references contained therein. Of course, a hybrid theory would not be ruled out.
12. M. Dummett (1954), 'Can an Effect Precede its Cause' Proceedings of the Aristotelian Society, Supplementary Volume 28, M. Dummett (1964), 'Bringing about the Past' Philosophical Review 73.
13. B. Russell (1912-13), 'On the Notion of Cause' Proceedings of the Aristotelian Society 13.
14. T. Honderich (1988): A Theory of Determinism (p.40-45)
15. B. Russell (1912-13), full reference in footnote 13.
16. W.V. Quine (1974), The Roots of Reference (Ch.1 s2); W.D. Hart (1988), The Engines of the Soul (Ch.5).
17. c.f. J. Fodor (1984), 'Making Mind Matter More', Philosophical Topics 67. Among differences between Fodor's approach and this one is that for Fodor the appeal to law states only a sufficient condition.
18. c.f. E. LePore and B. Loewer (1987), 'Mind Matters', Journal of Philosophy 84, for a discussion of a similar approach, and further motivation for it.
19. See Chapters 3 and 8.

Chapter 3

The Extent of the Challenge

The basic question that we shall try to answer in this chapter is: is there any way in which mental properties may be supervenient upon some physical properties which will render those mental properties causally epiphenomenal? The positive answer to this question, offered by some philosophers, has given rise to the thought that whether mental properties are physical or non-physical is really beside the point. We shall try to argue the opposite. Three ways will be considered in which mental properties might be physical and yet still causally epiphenomenal. In each case, we shall see that it is difficult to establish that the mental properties genuinely are causally epiphenomenal. Consequently we will have a justification for our own approach to the issue of epiphenomenalism.

(a) What is supervenience?

Supervenience is a relation between properties. It unfortunately has many different formulations.¹ First, we shall put forward a skeleton account of supervenience.

After that, we will attempt to delineate the main varieties in which it comes.

Suppose there are what we intuitively take to be two categories or families of properties.

One family of properties is supervenient upon another family of properties with respect to the members of a domain if and only if for any two members of that domain, or one member at two different times, if they are identical in their possession of members of the second family, then they are identical in their possession of members of the first family.

There are two variables to play with in the account of supervenience we have just provided. First, there is the extent of the domain. Second, there is the matter of how the underlined conditional should be interpreted.

When we speak of the domain of entities for which the supervenience relation holds we are thinking of those entities for which the conditional is said to be true. It allows various moderations of the strength of the claim that there is a link between two properties. However, the modifications which are offered are of doubtful utility if what we are genuinely interested in is a connection between properties. What sense can we give to such a relation if it is not between properties per se, but rather between two types of properties, only if they are possessed by certain individuals? Another way in which

the claims of supervenience may be moderated is by interpreting the conditional underlined as asserting a material implication, a relation of natural necessity, or a relation of logical or metaphysical necessity. For the latter two options we would add 'necessarily' in front of the conditional, indicating what sort of necessity we envisaged. Hereafter, we shall not bother to distinguish logical from metaphysical necessity.

(b) On the correct way to formulate the supervenience of the mental on the physical

We saw that supervenience was a relation between two families of properties. To formulate the supervenience relation for mental properties, we should take the first family to be, obviously enough, mental properties and the second family we shall call 'relevant non-mental physical properties'. 'Non-mental' because it is trivial that a property supervenes on itself, 'relevant' because nobody would claim that mental properties supervene on the entire family of physical properties but only some, those we have dubbed 'relevant'. Naturally, we should not stop at describing these properties as 'relevant'. Any body who adopts an account of supervenience is committed to there being a way of describing non-mental physical properties that does not just identify them as the ones upon which mental properties supervene. However, for the discussion

that follows we do not need to engage in such a specification. We are talking about the character of supervenience. There are other complications into which we also do not need to go.²

Donald Davidson³ is often taken to believe that the correct way to formulate the supervenience of mental properties upon the relevant non-mental physical properties is by taking the conditional to be just a material conditional. In which case, it is allowed that if all the members of the domain in question do not differ in their mental properties then they may have any relevant non-mental physical properties you like, since the universally quantified material conditional would be true. This dramatically highlights what seems unsatisfactory about the formulation, given that the account is the full story of the relation between these families of properties. It allows that the association of mental properties with the relevant non-mental physical properties that has occurred, has occurred, more or less, for no reason, or by accident. We shall not consider such an account further, although many of the conclusions to which we come apply to it as well.

The second suggestion as to how we might interpret the conditional in the account of supervenience originally provided, is that it captures a relationship of natural necessity. If such a relationship holds, then it is clear

that mental properties are not causally epiphenomenal since the relationship in question is causal. Of course, it might be argued that the occurrence of mental properties would be mere effects of the occurrence of physical properties, the latter being causes. However, it is certainly not guaranteed from the description of the relationship, that this is so. Moreover, the distinction between cause and effect is one about which we have already found reason to be uncertain. Consequently, we shall not consider this case further, however plausible it may be as an account of the relationship between mental properties and physical properties. We shall return to discuss it in the conclusion.

The third suggestion is that the conditional asserts a relationship of metaphysical necessity. Thus, it is said that for any two members of a domain, or one member at different times, it is metaphysically necessary that if they are identical in their possession of physical properties, then they are identical in their possession of mental properties. It is this interpretation that we shall primarily try to assess for its repercussions upon the issue of epiphenomenalism and it is what most have in mind when they talk of supervenience and its potential threat. Metaphysical necessity is taken to be a stronger notion of necessity than that which we called natural necessity, but this strength is not always taken to imply

that anything which holds of metaphysical necessity holds of natural necessity. So there is no quick way to deal with the present suggestion.

Over what domain of individuals should we take the supervenience relation to be defined? This is the second matter that skeleton account of supervenience invites us to resolve. Often it is said that although mental properties may be supervenient upon relevant non-mental physical properties for some individuals, in other possible worlds mental properties may supervene on some other properties than these. If people were composed, perhaps only in part, from very different non-physical non-mental stuff, mental properties may be supervenient upon relevant properties of that stuff. We should not want to say, in the case of those individuals, that if they are identical in relevant non-mental physical properties, then they will be identical in mental properties. So, we should limit the domain of individuals which we are interested in to those which possess both mental and only the relevant non-mental physical properties. Also, we might want to allow that mental properties are not essentially supervenient. In which case, there could be individuals with non-supervening mental properties.

Rather than attempt to delineate the precise domain of individuals about which we wish to talk, we shall take

the remarks above to highlight what we have already noted, namely, that to talk of a domain of individuals over which the supervenience relation may be defined is to be unnecessarily complicated. We want to focus on a relation between properties regardless of the individuals which possess them. With ^{UK}his in mind, the following seems to fit the bill.

Metaphysically necessarily, if an object possesses one of a range of non-mental physical properties $p_1, p_2, p_3 \dots p_n$ then that object possesses a certain mental property m_1 .

Such a conditional should be provided for each supervening mental property. It is this relationship, we may argue, that represents the truth behind the various supervenience claims and may reasonably be called a relation of supervenience itself.⁴

(c) **The relation of supervenience does not imply that there are causally epiphenomenal properties.**

Our aim will be to try to understand why somebody would think that the relationship of supervenience alone has repercussions for the causal efficacy of mental properties. What we shall try to show is that if we assume either side of the supervenience relation is causally efficacious, roughly the same conclusions follow as to the efficacy of the other side. Therefore, being in

a relation of supervenience is not a particular threat to the causal efficacy of either side.

Suppose first that the occurrence of a relevant non-mental physical property is said to be the cause of something else. Moreover, for ease of discussion, suppose we are talking of neighbouring links in the causal chain. Would the mental property that supervenes upon it be counted as causally efficacious? The question divides into two, depending upon which way of resolving the problem of spurious causation is adopted. According to the first option, the mental property would be counted as causally efficacious. We would be prepared to assert 'If the mental property had not occurred, then the effect would not have occurred', so condition (a) is satisfied. It might be thought that the alternative resolution of the problem of spurious causation provides a greater threat. The counterfactual conditional mentioning the occurrence of the mental property and the effect only holds, it may be urged, because the following two counterfactuals hold: first, that between the occurrence of the relevant non-mental physical property and the effect; second, that between the occurrence of the non-mental physical property and the occurrence of the mental property. However, it is at this point that we need to be told more about why the relation of supervenience holds. The solution to spurious causation was built to exclude an event being counted as a

cause of another, when that first event was distinct from and caused by one of the actual causes of that other event. It is not obvious that the same solution can be used to exclude the occurrence of the mental property from being causally efficacious since we do not know that it is distinct in the right way from the occurrence of the relevant non-mental physical property that is the cause. After all, a counterfactual may be asserted concerning the occurrence of a property and itself. Yet, we would not take this to rule out the causal efficacy of the property, since nothing would then end up causally efficacious. Thus, the threat mentioned cannot be considered exactly looming.

Suppose now that the occurrence of the mental property is causally efficacious. In those circumstances, what can we say about the occurrence of the relevant non-mental physical property? Again our discussion splits into two, depending upon the resolution adopted of the problem of spurious causation. On the first option, the occurrence of the non-mental physical property should be counted as causally efficacious. If we are allowed, as seems reasonable, to suppose that in the circumstances in which the mental property occurs, it could not have supervened upon the occurrence of another physical property then since the occurrence of the non-mental physical property was necessary to the occurrence of the

mental property, the effect would not have occurred if the non-mental physical property had not. If we are not allowed to suppose that the mental property is so limited in this case, it turns out that the relevant non-mental physical property is not a cause. Thus, strangely enough if the supervenience relation threatens the causal efficacy of anything it is the supervened upon properties.

Turning to the second way in which we may resolve the problem of spurious causation, it seems that the same qualms we raised with regard to the causal efficacy of mental properties a moment ago, would arise for the non-mental physical properties. So, it looks as if things are, more or less, on a par with regard to the threat the supervenience relation poses to the properties in this relation, which is what we were trying to show. Clearly, whatever threat is seen to arise from supervenience, must arise because of some further commitments those who assert such a relation have.

(d) **On the relation of macro-properties to micro-properties and the threat of epiphenomenalism which is said to ensue.**

A macro-property is one which occurs as a result of the occurrence of the micro-properties upon which it supervenes. Here are some examples. The macro-property of being a liquid occurs because of the properties that

the constituent particles of the liquid possesses. The macro-property of being a railway network occurs because of the properties that components of the network possess.

The notion of a macro-property should be compared with two other notions, that of a determinable property and that of a second-order property. A determinable property is the property which necessarily co-occurs with each member of a certain category of properties as a result of which those properties belong to that category. For example, the property of being coloured is a determinable property, the property of being a specific colour is a determinate of this determinable. Thus, one might say that the property of being yellow ochre is a determinate property. Only 'might' because it is plausible to claim that there are various shades of yellow ochre in which case the property of being yellow ochre is also a determinable property, and so on. Of course, if one views the determinable-determinate distinction as a relative one, so that a property can be a determinate property relative to some determinable property, yet a determinable property relative to some other determinate property, then we could happily conclude that the property of being yellow ochre is determinate. We should not try to resolve this matter here.

A determinable property need not be a macro-property because whereas the latter, by definition, requires for

its occurrence more than one property to occur even if this is not true of the properties upon which it supervenes, determinable properties do not require this unless its determinates do. The occurrence of a determinable property such as colour only requires the occurrence of other properties such as having extension because its determinate properties do.

A second-order property is a property possessed by properties. The property of being a colour property is possessed by, for example, the property of being yellow and hence is a second-order property. The difference between macro-properties and second-order properties is obvious. Although it is true that macro-properties can only occur if the micro-properties upon which they supervene occur, it is not true that all macro-properties are properties of other properties.

For the rest of this section, we shall only be interested in macro-properties and their relationship to micro-properties. More particularly, we will examine the claim that causation at the micro-level robs macro-properties of their causal efficacy.⁵

We have already found that the relation of supervenience does not detract from the causal efficacy of either of the relata. Therefore, if the macro-micro relation robs the macro-level of causal efficacy, it must be because of particular features that the relation

possesses. The most obvious feature is that the occurrence of macro-properties required the occurrence of two or more micro-properties. Someone may reason that

- (1) What causally determines the occurrence of the relevant micro-properties thereby determines the occurrence of the relevant macro-properties.
- (2) The occurrence of micro-properties alone causally determine the occurrence of micro-properties.

Therefore:

- (3) The occurrence of micro-properties alone causally determine the occurrence of macro-properties.

It is only a short step from here to the claim macro-properties are causally epiphenomenal, as most if not all properties can be fitted into the macro-micro division. The first premiss of the argument is clearly true. It is the second which causes the trouble.

Suppose someone tears a page from corner to corner, with the result that there are now two triangles rather than one rectangular page. We might initially describe the causal relationship by saying that the event of being diagonally torn was caused by the event of human hands gripping the page in a certain way and a certain force being exerted. The properties that we have used to describe these events are clearly macro-properties.

Intuitively, we might think we had a case of macro-causation but, remembering the position of those who say that macro-causation is not real, but an illusion derived from micro-activity, we stop ourselves from jumping to conclusions, and take a closer look.

No doubt there is a micro-story to tell about the tearing of the paper. What is in doubt is whether it exists at the expense of macro-causation. When the page is torn, the molecular bonds which constitute the page are broken. The activity of the hands breaks the bonds. So, why is it not reasonable to conclude that the occurrence of macro-properties does have an effect upon the occurrence of micro-properties?

Perhaps the idea is that we should, out of parsimony, take the causal connections between the occurrences of the micro-properties as determining the causal connection between the occurrences of the macro-properties. Suppose it is the case that for the occurrences of two macro-properties thought to be causally related, we can show that each occurrence of a micro-property upon which the occurrence of the macro-property constituting the effect supervenes is causally related to an occurrence of a micro-property upon which the occurrence of the macro-property constituting the cause supervenes. Might we not then say that the causal relationship between the occurrence of the macro-properties should count as no more

than the sum of the causal action of the occurrence of the relevant micro-properties? Maybe one should say this, but from making this claim it does not follow that there is no causal relationship between the occurrence of the macro-properties. It would be no more plausible to hold this, than to hold that since the occurrence of macro-properties is no more than the joint occurrence of the micro-properties upon which they supervene, the macro-properties do not exist. Surely the obvious thing to say is that just as macro-properties supervene upon micro-properties; so do macro-causal relationships supervene on micro-causal relationships.

It is at this point that it might be felt we should introduce into our discussion something which has been long overdue, laws of nature. It will be suggested that it is by considering which properties figure in laws of nature that we shall discover what is causally efficacious and what appears so, but is in fact causally epiphenomenal.

It is questionable whether an appeal to the notion of laws of nature will be successful in establishing that macro-properties are causally epiphenomenal. The first thing to note is that we provided an account of causal relationships which made no appeal to laws. So, it would have to be independently established that a property which satisfied the account but did not figure in a law of

nature should be considered causally epiphenomenal. Second, some sense will have to be made of what it is to figure in a law of nature, and this is by no means clear cut.⁶ However, since any problems on this front will favour the position that we are urging, and work against the opposition, let us take it that such a notion is intuitively clear and see where that gets us.

The thought might be that the existence of laws explains, in some sense, the existence of causal relationships between occurrences of properties. If this is true then it may be thought that the type of law in which they figure reveals something about their causal efficacy. In what follows, we shall consider whether there are any grounds for thinking that certain laws do reveal something of this kind.

One way in which something may fail to be causally efficacious is if it figures in what we may call a 'derivative' law, a law which has been deduced from some more general law. Here is an all too simple illustration of the idea. Suppose the following law is true: 'Under the conditions of atmospheric pressure generally found at sea level on earth, all water boils at 100°C'. Obviously, this law, itself, may be derived from some other more fundamental law, but this does not matter for our purposes at hand. Suppose, further, that we are interested in the water in the mountain spring

purportedly used by Perrier, and we put forward the following law: 'Under the conditions of atmospheric pressure generally found at sea level on earth, all the spring water used by Perrier boils at 100°C'. The second law is less general than the first. It bespeaks of a sub-type of water. For the purposes of illustration, it does not particularly matter that the sub-type is of no scientific interest. What we should note is that it is plausible to think that the second law holds only because the first law does and the spring water in question is water. We may derive the second law from the first. So, it seems reasonable to say, whatever causal efficacy the second law points to, is no more than the causal efficacy the first law records.

Parsimony once again comes in, but this time it is not open to the criticism we made before. We are not taking the causal relationship specified by the derivative law as non-existent. Instead we are saying that we can view the causal relationship between spring water at 100°C and boiling, as the very same causal relationship as that between water at 100°C and boiling. The property of being from a certain spring is shown to be epiphenomenal.

More generally, we can say that:

A causal relationship holds between two entities as a consequence of the entities possessing or being composed from the occurrence of particular properties

if and only if the causal relationship between these entities only holds because

- (a) The occurrences of these properties satisfy the account of causal relationships we gave previously, for the relationship in question.
- (b) The properties in question are the most general, or highest in the hierarchy of determinables to determinants, for which condition (a) is satisfied.

The account is formulated in these terms, rather than directly in terms of properties to hark back to our discussion in the previous chapter. The property of being spring water at 100°C is lower in the hierarchy of determinables to determinates the property of being water at 100°C.

It is by no means obvious that micro-properties will figure in non-derivative laws, rather than macro-properties. If a macro-property is constantly conjoined with another macro-property, and each of these macro-properties supervenes upon a number of different collections of micro-properties, then if a causal relationship holds, it holds between the macro-properties rather than between the various collections of micro-properties upon which they supervene. The macro-properties are more general than the collections of micro-properties which variously occur.

Now, it might be said that this cannot be so. Physics holds out the promise of causally explaining all that goes on in terms of the properties of a few, fundamental constituents of the universe. This must be more general than the causal explanation of the occurrence of one macro-property in terms of, say, the occurrence of another macro-property.⁷ Of course, this is true, but it does not touch the point we were making. We dealt with this matter in our opening remarks concerning how macro-causation should be taken as constituted from micro-causation. The point that we were trying to make concerned the causal relationship that held between two entities and not just that which held between the constituent parts of the entities. We said that, for the causal relationship between the entities, the causally efficacious properties were those highest up in the hierarchy of determinables to determinates which satisfied the account of causal relationships in general.

A second type of law which might be thought to imply that the properties which figure in it are not causally efficacious are those laws which involve qualifying conditions. For example, the law may be 'Under circumstances C, all events of type F are followed by events of type G' or 'All things being equal, all events of type F are followed by events of type G. The phrase 'all things being equal', or 'ceteris paribus', is a

special case of a law involving qualifying conditions where, for various reasons, the background is left inexplicit, the idea being that if nothing is out of the ordinary the succession will occur. It has been suggested that what is peculiar to one science, physics, is the absence of qualifying conditions in its laws. Suppose that there were such a science, whether or not it is physics, and that for all other sciences there were such conditions. Would anything follow with regard to the causal efficacy of the properties identified by the various sciences?

In our discussion of causality, we said that what we called 'a cause' was generally one member of a collection of events which are sufficient for the effect, or the determination of the probability of the occurrence of the effect, regardless of the circumstances in which they occur, unless, of course, one counts as part of the circumstances the laws of nature which hold. One might then choose to re-write this relationship by specifying a particular type of circumstances by listing all the types of members of the collection of causes bar one, an event of type F, and saying that in circumstances of that type, all events of type F will be followed by events of type G'. Since we would have, by rewriting a causal relationship in this way, one instance of the general form of the law 'Under circumstances C, events of type F are

followed by events of type G', we may conclude that, just because a law has this form, it does not follow that the properties mentioned in it lack causal efficacy. If there is a threat it must come from elsewhere. Even the laws of the physics can have its laws rewritten in the damning fashion.

The situation may be thought to be somewhat different when we turn to 'ceteris paribus' laws. For such laws, there are, in all likelihood, a number of circumstances in which events of type F will be followed by events of type G, the succession of the events depending not only upon these circumstances but also on the nature of the properties upon which the property of being F supervenes. For example, consider the following law concerning writing implements: all things being equal, all applications of a writing implement to a surface in a certain manner (which would obviously need to be specified) will result in things being written. There are, of course, various circumstances in which this succession of events occurs, since different writing implements make marks on different surfaces, such as paper, a white board, a blackboard and so on. Moreover, it is not the case that in each of these circumstances the succession of events occurs. If the writing implement is chalk, for something to be written, it must have been applied to a blackboard; if a pen, then it must contain ink and be applied to a piece of

paper, and so on. The succession of events depends not just upon the circumstances but upon what the macro-property, of being an applied writing implement, supervenes.

The main reason for denying causal efficacy in the case just described may be that the occurrence of the causal relationship seems to depend upon features of the supervened upon properties, and not merely upon features of the supervening macro-properties. So, one is apt to think that the macro-properties do no work at all. Whatever causal efficacy there is rests upon what is required in detail for the required effect to occur, namely, the relevant micro-properties. However, it is hard to see why one should not include a specification of the supervening properties among the collection of causes, which would also include the relevant supervened upon properties, that are jointly sufficient for the effect. To assume that the supervening property is superfluous, when it would otherwise be judged, by the account we have offered, to be causally efficacious, is to give up the hope that we might find some independent reason for believing that properties which figure in a certain type of law lack causal efficacy. It might be said that the reason why the macro-property should not be included is that there are times when it occurs, and the effect does not. However, this situation holds for most

properties we are quite willing to count as causes. All 'causes' only cause things in certain circumstances.

One feature of the laws of some sciences is that if the circumstances were specified, in which two types of events co-occur, the specification would be in terms of properties not part of the subject matter of the science. So, it might be decided that one should drop all mention of these circumstances and use a phrase such as 'all things being equal'. The decision has the upshot that the causal efficacy of the properties is not made manifest by showing how they relate to the conditions in which they bring about the occurrence of other properties, however such statements presumably have other virtues. What we have failed to find not only in this case, but quite generally, is an argument for the putative unreality of macro-causation.

(e) Functionalism, dispositions, powers, capacities
and the threat of epiphenomenalism

It has been argued recently⁸ that if Functionalism was correct, mental properties would be causally epiphenomenal. It is this claim that we shall try to assess in the present section.

Suppose we think of a mind as a certain sort of system with inputs in the form of stimulation and outputs in the form of behaviour, or attempts to behave in a

certain way, if, for example, a subject, is paralysed. Functionalism, as an all encompassing theory of mind, holds that mental events are events which possess a definitive role in the system, with respect to other mental events, stimulation and behaviour. If one takes Functionalism to be a complete theory of mind, then each type of mental event is supposed to be so characterised. One may, of course, have functionalist theories of only specific mental events, such as beliefs.

Two things were left deliberately underspecified in our characterisation of functionalism. First, there is the way in which mental events are related to the other things mentioned. There are two options here. One is to allow that there is a distinct type of 'functional' relation. A mental event, by having functional relations to many things, has a functional role. These functional relations are understood to be the sort of things we specify when we describe how a certain type of machine works. Various parts of the machine are identified and the role these parts play in its overall functioning specified. For example, one part of the bicycle enables the bicyclist to turn the wheels, namely the pedals and chain. If we were to specify this part's functional role we would say that as input it received pedal pushing and as output it supplied a certain force at the hub of the back wheel, the input being correlated with the output.

It is natural to hear the specification of the functional role of some component of a machine as the description of the causal relations it has with other components, or its 'potential' causal relations. Nevertheless, although it is natural to hear the functional specification in this way, the thought that a functional specification could only be of the causal operations of some system is supplementary to the idea of a functional specification per se. In the latter, no claim is made as to how a machine may be, apart from that which is a straightforward implication of the specification. With this in mind, we are in a position to recognise the second option open to the proponent of functionalism. He or she may say that minds are just those types of system which involve specific causal operations. In which case, the functional specification offered of a mind will be a description of the causal role of mental events. For each mental event we specify a distinctive causal role, the causal relations it would have in various situations.

As far as our present discussion is concerned, it does not matter which option we adopt. For a mental event to be causally efficacious it must be part of a mind comprised of causal operations, so we shall henceforth talk of the causal role of mental events regardless of the niceties just mentioned.

The second way in which we underspecified the nature of functionalism was by failing to articulate what we meant by saying that mental events were 'characterised' by their causal role. Here are three suggestions. First, a type of mental event is defined to be that which in fact has a certain causal role. If this were the correct interpretation, then the proponent of functionalism would be merely making a suggestion as to how we may identify mental events. We just are to look for things which play a certain causal role. It is not that they must play such a role. A parallel case would be to say that a cow is the type of thing we find attached to milking machines in dairy farms. There is no reason to believe that they must be so attached, but for someone who did not know what a cow was, this fact would be helpful.

A second suggestion is that types of mental events are those which necessarily play certain causal roles. An entity which ceased to play the role would cease to be a type of mental event. However, the event is not just to be specified as the occurrence of the macro-property of playing a certain causal role. Instead, it is that thing whatever it is, which so plays it. To return to the example we used before. We may define a dairy cow to be a cow attached to a milking machine in a dairy. In which case, when the cow ceases to be attached to the machine, it no longer is a dairy cow. However, it is not just the

fact of attachment that makes a dairy cow a dairy cow. We also need to pay attention to what is attached.

The third and final suggestion is that what we take to be mental events are just the occurrences of macro-properties of playing certain causal roles. Unlike the second option, the putative mental event is not seen as in any way composed from the thing which plays the causal role. It is specifically this suggestion that can seem to raise the challenge of causal epiphenomenalism as far as the argument of this section is concerned. So, it is worth pausing for a moment to see which of the three suggestions most captures what the proponents of functionalism wish to say.

The first suggestion was merely put forward to provide a contrast with the second and third claims. The proponent of functionalism, of whatever sort, does not merely want to assert that we may in fact use the causal role of mental events to identify them, he or she wants to say something about their nature.

The second suggestion was put forward by David Lewis and David Armstrong.⁹ A mental event, according to Lewis' account, will have properties other than the property of playing a certain causal role. Now, a question arises as to the status of these other properties. Are some of them to be taken to be mental properties or should we only count the property of playing a certain causal role as a

mental property? It is reasonable to think that the proponent of functionalism as a complete theory of mind will take only the latter type of property as mental. In which case, the argument that follows has application to this account as well.

Before we consider the argument that has been put forward in favour of the causal epiphenomenalism of mental properties if mental properties turn out to be the properties defined by functionalism (hereafter, 'functionally defined properties'), it is worth noting the extent of the application of the argument. Functionally defined properties may be plausibly viewed as just a special case of those properties that have been called 'powers', 'dispositions' and 'capacities'. There is one complication. Just as there is a distinction between those who hold that mental events are the occupants of a particular causal role, rather than just the occurrence of the causal role, so there are those¹⁰ who hold that occurrences of the properties just mentioned, powers, dispositions and capacities, are the occupants of the causal role associated with them rather than the occurrence of that causal role. Thus two interpretations may be given of the claim that, for instance, the power to burn things is that property an entity has if in certain circumstances, it would set something alight. Upon one view, the power is that property of the entity that

enables it to have the effect of setting something alight. Upon the other view, the power is just the property of setting things alight in certain circumstances. It is the latter understanding of powers and the rest which is analogous to functionally defined properties. Their common feature is that their whole character is specified and, thereby, principally defined in terms of what would happen in certain circumstances. In other words, their nature is given in terms of possible effects. It is this feature we shall focus on.

The argument against the causal efficacy of these properties is basically this. Occurrences of the properties just mentioned cannot be the causes of the effects in terms of which their nature is defined, the 'defining effects'. Equally, they cannot be the causes of any effects of the effects because they are not causes of the defining effects. Consequently, the properties have no causal efficacy.

The claim that the occurrence of the properties with which we are concerned cannot be causes of the effects of the effects should not be controversial given that they are not causes of the defining effects. By definition, the defining effects are the immediate causes of their own effects. Therefore, the only way the properties mentioned could, by their occurrence, be causes, is through being further back along the causal chain. But this is what has

been ruled out by saying the occurrences of the properties in question do not cause the defining effects. Clearly, it is the latter claim that will excite controversy.

The argument for the claim is probably this. For the occurrence of one property to cause the occurrence of another property, there must be something in the former's character which brings about the occurrence of the latter. But the character of the properties in question is entirely given by the statement that if they are possessed in certain circumstances, certain other properties will occur. Thus, the character just attributed to these types of properties does not enlighten us as to what it is about their nature or character which is causally responsible for the effect occurring. In fact, there is no such feature of their nature allowed. So, one wants to say, there should be some other property which explains why the effects occur and thereby also accounts for the possession by an entity of the properties we are considering. This line of thought seems particularly threatening for functionally defined properties as it is already accepted, by those who propose they exist, that they depend for their existence upon the occurrence of other properties that are causally related to each other.

An example or two might help to bring out the point just made. Suppose we are interested in providing a causal explanation of why sleeping tablets send us to

sleep. We are surely not helped if we are told that it is because they have the power to send us to sleep. Of course, if they do have such a power, it follows that they will send us to sleep. But, to mention the power is just to mention that the causal relationship in fact holds. It does not state why it holds. Consider, now, the functionally defined property that might characterise the sensation of thirst. Suppose for simplicity's sake that the following is partly definitive of the property: if someone were to have the sensation of thirst and were to believe that he or she is near a drinking tap, then he or she would drink from the tap. Given this definition we surely cannot suppose that we may give an explanation of a person drinking in terms of his or her sensation of thirst since the latter just has its character defined by the fact that the person would drink in the relevant circumstances.

On the other hand, our account of causal relations commits us to allowing that functionally defined properties, and the others, are causally efficacious. If we ask 'Would the person have drunk from the tap if he or she had not had the sensation of thirst?' The answer is no. Consequently, the sensation of thirst is necessary in the circumstances, and satisfies our test for causal efficacy.

A possible resolution of the present conflict is not

hard to find. Functionally defined properties, and the others, are shown to be causally efficacious with regard to certain effects because they are defined to be those things which are only possessed by entities if the entities give rise to the effects in question, in the relevant circumstances. As a result, they are deliberately so defined as to free ride on the causal efficacy of others. This suggests the following supplement to our account of what it is for a causal relationship to hold between two things as a consequence of properties they possess.

- (c) The properties do not satisfy our account of causal relationships just because one of the properties is specified to occur only if the other property occurs in certain circumstances, and the first property is said to have no other character than that which has just been specified.

The addition to our account would rule out functionally defined properties, and other properties understood in the same way, from being causally efficacious.

Although the principle we have just put forward is intuitive, it is difficult to feel secure that it is true. One reason for the feeling of insecurity is that some philosophers¹¹ have argued that the world is only composed of properties that are defined in terms of their possible

effects, yet they do not conclude from this that there is no causality in the world. Our principle would yield such a conclusion.

Fortunately, we can turn insecurity to advantage. If we had chosen to establish the causal epiphenomenalism of mental properties by the means we have just canvassed, we would have to, first, defend the claim that mental properties are defined in terms of their possible effects and, second, defend the principle, we have put forward, against the alternative approach to causal efficacy outlined by those philosophers who have argued that the world both contains causation and only properties defined in terms of their possible effects. This would be a tall order. It would have been considerably more difficult to come to some conclusion about whether mental events have some role to play in psychological explanation had we adopted this approach, than by adopting the approach we have. We, therefore, have some justification for the way we have proceeded.

The reasoning that leads up to the claim that mental properties are functionally defined properties is far less compelling than the argument that mental properties are non-physical, as the virtually continual onslaught on such a claim bears witness.¹² Many, one suspects, would be inclined to view causal epiphenomenalism as the last nail in the coffin of any such view. So, dialectically, it

would not have been wise to consider such an argument in contrast to our own.

The defence we offered, of the principle, is also likely to come under pressure. The principle stated that a property should not be taken as a cause of those effects whose occurrence in certain circumstances is a condition of the occurrence of the property. The defence was that it would not be illuminating from an explanatory point of view to cite such a property. In reply, it may be argued that the absence of explanatory illumination is not sufficient to establish that the thing cited as a cause is not a cause. Explanation, it may be said, comes apart from causation. What makes the citation of one such property explanatory, where another would not be, is that the first property is a component of the collection of properties whose joint occurrence constitutes an entity's possession of a complex property defined in terms of possible effects. The citation of the component property is explanatory of both the occurrence of the complex property and the possible effects definitive of the complex property. In this way one may account for the intuition we have that we cannot explain why a sleeping pill makes us sleep by saying that it has the power to do so. The attempted explanation in this case is not explanatory because we have only cited the power whose nature is characterised by the effect it has, and not a

component power of that power. Our recognition later on that causal explanation does not exhaust the types of explanation there are strengthens such an attack.

In the present chapter, we discussed three ways in which it might be thought that mental properties would be causally epiphenomenal even if they were physical. We first considered the case in which mental properties were quite generally supervenient on some physical properties. Nothing was seen to follow regarding whether or not they were causally efficacious. We then wondered whether a particular type of supervenience had such an implication, so we considered that of macro-micro supervenience. Again, nothing was seen to follow. Finally we considered the type of supervenience one would assert if one thought that mental properties were functionally defined properties. Our conclusion was that there was a case to be made here, but that it depended upon too many debatable premises for it to threaten the rationale of our own inquiry, which is to consider the most compelling arguments for epiphenomenalism. With these points in mind, it seems there is some reason to persevere with the focus that we have adopted.

References

1. c.f.J. Kim (1978), 'Supervenience and Nomological Incommensurables' American Philosophical Quarterly 15, J. Kim (1984), 'Concepts of Supervenience'

- Philosophy and Phenomenological Research 45, J. Kim (1987), "Strong" and Global" Supervenience Revisited' Philosophy and Phenomenological Research 48; J. Bacon (1986), 'Supervenience, Necessary Coextension and Reducibility' Philosophical Studies 44.
2. See references in footnote 1.
 3. c.f. D. Davidson (1985), 'Reply to Harry Lewis' (in B. Vermazen and M.B. Hintikka, eds., Essays on Davidson, Actions and Events). Also, his claim that any generalisation concerning mental properties and physical properties will not be counterfactual supporting only holds if we interpret supervenience this way.
 4. c.f. S. Blackburn (1985), 'Supervenience Revisited' I. Hacking, ed, Exercises in Analysis. The formulation also avoids the problems connected with the difference between 'strong' and 'weak' supervenience
 5. A claim seemingly put forward in J. Kim (1979), 'Causality, Identity and Supervenience in the Mind-Body Problem', Midwest Studies in Philosophy IV, J. Kim (1984), 'Epiphenomenal and Supervenient Causation', Midwest Studies in Philosophy IX, (1989), 'The Myth of Non-reductive Materialism', Proceedings of the American Philosophical Association Vol. 63. It is hard to be sure whether Kim is a 'reductionist' or 'eliminativist' about macro-causation but the thrust of his remarks in the 1984 paper and his categorisation of supervenient causation as a species of epiphenomenal causation suggest the latter.
 6. e.g. D.M. Armstrong (1988) What is a Law of Nature?
 7. I am grateful to David Ruben for pointing this out.
 8. N. Block (1987), 'Can the mind change the world?' (in G. Boolos (1990), ed., Meaning and Method).
 9. D. Lewis (1966), 'An Argument for the Identity Theory' Journal of Philosophy 63 (both reprinted in his (1983), Philosophical Papers Vol.1.), D. Lewis (1970), 'How to Define Theoretical Terms' Journal of Philosophy 67, D. Lewis (1972), 'Psychophysical and Theoretical Identification' Australian Journal of Philosophy 50; D.M. Armstrong (1968), A Materialist Theory of Mind (Ch.6).
 10. See D.M. Armstrong (1968), reference in footnote 9.
 11. H. Mellor (1976), 'In Defense of Dispositions' Philosophical Review 83; S. Shoemaker (1980), 'Causality and Properties' (P. van Inwagen, ed., Time and Cause).
 12. e.g. N. Block (1978), 'Troubles with Functionalism' (in N. Block (ed.) (1980) Readings in the Philosophy of Psychology Vol.1), (1980) 'Are Absent Qualia Impossible?' Philosophical Review 89.

Chapter 4

Two ways in which mental events might be non-physical

How may mental events be non-physical? One way would be if they lacked spatio-temporal location. However, this looks unlikely to be true. The other way is if their existence implied that a subject was aware of them. It is this possibility that we shall examine. Instead of talking of parts of the nature of mental events we will talk of properties which belong to their nature. Our aim, therefore, will be to find out whether any non-physical properties so belong.

In the present chapter, we are going to describe two properties that may be considered peculiarly mental. They have been selected because it is often these properties that are taken to be non-physical. If the threat of epiphenomenalism has application, the case for it starts here. Both of the properties relate to the content of mental events, their phenomenal content and intentional or representational content respectively. Nothing in this chapter will constitute a defence of their existence, that task will remain for subsequent chapters, so although we

talk of their existence, we should not as yet be taken to be so committed.

(a) Phenomenology

Phenomenology may be defined to be the appearance of the mind to the subject whose mind it is. Obviously this notion is closely connected with that of consciousness. However, we shall not identify consciousness with phenomenology because consciousness may have features other than phenomenology. It would be better if we were not committed to ruling this out.

Phenomenology has also been characterised by some as what it is like to be a certain sort of thing. For instance, Thomas Nagel¹ has argued that there is something which is what it is like to be a bat. Whatever one's views of bats, it is reasonably clear that we do have some idea of what Nagel means. Of course, there are various ways in which one may understand the thought that there is something which it is like to be a creature of a certain sort. I may be like a bat in that I have wings and prefer to rest by hanging upside down. This is not the sense Nagel is interested in. What he says is this:²

'...fundamentally an organism has conscious mental states if and only if there is something that it is like to be that organism - something it is like for the organism'.

The addition of 'for the organism' is crucial. The idea presumably is that there is a way in which one can be like an organism of which the organism is aware. As, the notion of recognition or awareness appealed to here is obviously one which involves consciousness, 'what it is like to be an organism of a certain sort' is to be interpreted as what it is like to be the organism of which the organism is conscious.

Ted Honderich³ has properly wondered whether any illumination is shed by Nagel's attempted articulation of what consciousness is. The discussion of the previous paragraph, suggested by Honderich's remarks, serves to reinforce the latter's position. As soon as we see what is being smuggled into the 'something it is like for the organism' we recognise that little advance has been made. Nevertheless, some has. The phrase 'something it is like for the organism' may involve us bringing to bear our understanding of consciousness, but by appealing to our understanding of this phrase Nagel manages to single out one aspect of consciousness which he considers important.

Some time has been spent upon Nagel's characterisation because our account should be distinguished from it. We are not just concerned, as Nagel is, with those characteristics of mind which typify what it is like to be an organism of a certain type. For all we know, there may be no such general characteristics.

What we are interested in, to repeat, is the appearance of the mind to the subject whose mind it is, regardless.

To those who hold that it is illegitimate to say that the mind has an appearance, we have some defence. First, the phrase 'the appearance of the mind' just means the content of that range of experiences we have which are experiences of what we take to be our minds. To honestly doubt that we have such experiences is to render more than just slightly mysterious our inclination to believe that we do have some sort of mental life. Even those who wish to suggest that the idea of the 'appearance of the mind' involves a false picture will have to explain the ease with which the picture is adopted and purveyed, and this, it seems reasonable to maintain, will take us back to the claim that there are certain experiences we take to be of the mind. We may appease those who wish to view philosophers as unwitting cartoonists purveying caricatures of the world if we talk of the 'appearance of the mind', but admit that how we interpret this phrase is open to question.⁴

The concessionary stance we have just adopted would be ruined if we then made the assumption that the phenomenology of mind was composed from properties possessed by mental events. For, in making the assumption, we would have taken back exactly what we promised to adopt in the previous paragraph, namely the view that we should

be neutral about the interpretation of the phrase 'appearance of mind'. The battle between those who doubt that such a notion makes sense and those who are sure such a notion makes sense was construed as a battle over how to interpret this phrase. So, instead, we shall say that one way in which mental events may be non-physical is by possessing properties that constitute the phenomenology of our mental life should these properties exist.

(b) Intentionality

Most modern discussion of intentionality credits Brentano for first emphasising its importance, although, as we shall see, he in turn refers us back to the scholastics of the Middle Ages. The claim is that all mental events have as a constitutive part of their nature an act of presentation or representation. In the following passage, he develops the idea.⁵

'Every mental phenomena is characterised by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomena includes something as object within itself, although they do not all do so in the same way. In presentation something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired, and so on'.

The passage from Brentano has been quoted at length so that we may be clear as to the source of modern discussions of intentionality.

Brentano's notion of intentionality is mainly concerned with representation and not with what we have called phenomenology, or, thereby, mental content in general. However, contained within it are confusing strains of what we should wish to call phenomenology. We shall try to separate these from the basic idea which modern discussion has taken up.

Many mental events, the 'propositional attitudes' such as beliefs and desires, are sometimes ascribed to a subject using sentences with a special character. The special character is supposed to reveal the nature of an aspect of these mental events. The sentences contain a sub-clause, following the name of the relevant mental event and the subordinating conjunction 'that', which is itself a sentence, for example, John believes that it is raining. The description of the belief in this manner is one manifestation of what is surely a fact: beliefs are about things in the world. The sub-clause 'It is raining' expresses the proposition to which John has the attitude of belief and thus describes the mental event's 'representational' or 'intentional content'. We shall use these words 'representational' and 'intentional' interchangeably to indicate that type of mental content

that solely involves making reference to things. It is characteristic of the representational contents described by sentences expressing propositions that they may be assessed for truth or falsity. We shall call such representational contents, propositional contents. The intentional or representational content need not be a proposition. If John not only believes that it is raining, but also loves Nancy, there is one mental state John has, that of love, whose intentional content is not a proposition but Nancy.

As we have just seen, on the more modern understanding of intentionality, mental events possess intentional content either by making reference to items that putatively exist, or by being assessable as true or false, or indeed by satisfying both these conditions. The passage, on the other hand, suggests that the reference that is being made is to a 'content' or 'object' which, it is said, should not be understood as meaning a thing. Presumably the claim that the reference to an object should not be considered reference to a thing is another way of putting the point that mental events that have intentional content should not be taken to make reference to items in the world at all.

Why is it that Brentano should want to claim that our mental events do not make reference to items in the world under any circumstances? It can only be because he notes

that we can have a thought about such things as unicorns even though no unicorns exist and this thought is, in one crucial respect, just like a thought about pitbull terriers which certainly do exist. The thoughts are just like each other in that the existence or non-existence of the subject matter does not show up in thought. The phrase 'just like' indicates the infiltration of phenomenology into a discussion of intentionality. Otherwise, there seems no reason why we could not allow there to be a component of our mental events which makes reference to an item in the world if there is such an item, and just engages in the process of making reference without success if there is no such item. The fact that Brentano felt compelled to provide an object to which a subject could not fail to refer suggests that he was postulating objects of reference in response to the conclusions one would come to if one examined what it felt like to make reference on the occasions described.

Brentano may insist that we have misinterpreted what he meant. The notion of reference he was interested in has nothing to do with what a mental event concerns. He might say that the notion of 'reference to a content', or 'direction toward an object' is supposed to report a phenomenological fact, that fact being that we experience our mental life as a self in some way connected to the content of our mental lives. Ted Honderich⁶ has

discussed this matter in some detail, differentiating between various kinds of content. It is not a matter we need to go into here, because it does not in fact touch our present discussion. If that was Brentano's main point, then he is not the rightful source of the modern notion of intentionality in which we are interested. However, the basis for the interpretation of Brentano that we offered was not the phrases just mentioned but his talk of every mental phenomena containing within it something as an object which has 'intentional inexistence', as he so pellucidly puts it. The claim is that this notion should be understood to be connected with the notion of reference to things, that we described above.

Nevertheless, although the attention Brentano seemingly paid to the phenomenological character of the mental events was unwise, in this context, and resulted in him having to claim that the intentional content of mental events does not refer to items in the world, which is somewhat counter-intuitive, one should not suppose that avoiding the view he comes to is necessarily an easy matter. The difficulty is to understand how the world is represented as being in such and such a way regardless of whether the process of making reference succeeds in picking out anything in the world. Some aspect of our mental events surely makes it the case that a thought that unicorns have a horn may be correctly ascribed to a

thinking subject. It is tempting to suppose that that aspect of our mental events which makes the subject matter unicorns is one of the things, which are not items in the world, upon which subjects' thoughts are directed, especially, bearing in mind that the idea of making reference unsuccessfully cannot be identified with the aspect just described, because 'making reference' is a description of the activity of thinking and not what is thought.

To understand what may be wrong with this line of thought, it is first important to recognise a certain distinction, that between the means of representation and what is represented. Crudely, the difference between them is the difference between what we use to represent something, for example, a language, paint, or whatever, and what is said or expressed by what we use. Sometimes what is said or expressed is a proposition. Other times a question is asked, or a command given. We shall focus on those cases where what is said is a proposition.

When what is represented is true, it is tempting to consider that the very nature of what is represented is just some particular state of affairs. For instance, what is represented by the sentence 'John Major is the current British Prime Minister' may be thought to be the state of affairs described by that very sentence. It seems implausible to distinguish between what is represented and

a state of the world and characterise the nature of the former independently.

However, the situation changes when we consider those occasions where what is represented is false. If a false proposition is said to exist, then it cannot be the state of affairs that it represents as being the case since, because the proposition is false, there is no such state of affairs. Often philosophers at this point start talking about false propositions being such shadowy objects as non-actual states of affairs. A problem with this option is that one would still have to conclude that false propositions do not actually exist. In which case, it is puzzling how sentences and other means of representation may actually be false.

At this point, one might decide to bite the bullet and start talking of propositions or 'thoughts' being inhabitants of a third realm, or whatever.⁷ In which case, one would probably have to say that the means of representation was in some way related to a proposition in this realm whose character was to represent the world as being in a certain way. It naturally follows, upon this view, that how the world is represented as being may not be how it is.

Whether or not one does adopt this approach depends upon a host of issues that we cannot hope to discuss here. What we can do is recommend a way of conceiving of

intentional content which, at least, postpones our commitment to the existence of a third realm, and, also, perhaps points to the flaw in the reasoning we suggested lay behind Brentano's approach to intentionality. Instead, we shall say that the specification of what is represented is the specification of the 'intentional' capacity that a means of representation has. The capacity in question is to refer to the object or state of affairs, that we take an intentional content to concern, if the object or state of affairs exists. If the object or state of affairs does not exist, no reference is made. For example, the sentence 'John Major is Prime Minister' has, among other things, the capacity to refer to John Major if, should John Major exist, it refers to John Major. If John Major does not exist, then the capacity is unmanifested. The sentence 'Unicorns have a horn' would refer to unicorns, if they exist, but as it happens, does not so refer. True sentences are those which manifest their capacity to refer to the states of affairs that make them true. False sentences are those sentences that have not managed to manifest their intentional capacity.

A capacity, in itself, is not an activity of making reference. So, the line of thought we considered, in regard to Brentano's views, was right in stating that the idea of making reference alone could not be utilised to account for the nature of thoughts about non-existent

items. Where it went wrong was in suggesting that we needed to conceive of intentional contents as objects rather than capacities. Once one sees that intentional contents are capacities to make reference, the need to postulate the existence of intentional objects is ameliorated.

Although there are complications that we have passed over, the following general description of the nature of intentional properties (the constituents of intentional contents) will be provisionally adopted. Intentional properties, should they exist, are capacities or sub-capacities to refer. An intentional property refers to its referent only if the referent exists. In those cases when the referent does not exist, no reference is made, although we talk as if a reference has been made to characterise the capacity. We may perhaps take the notion of referring to capture what is common in both cases. 'Unicorn' is at all times referring to unicorns in spite of the fact that one assumes there is not a single successful case of reference.

We shall close the present chapter by making two points of importance with regard to the argument in subsequent chapters. First, one should avoid confusing

intentionality with phenomenology, and mental content in general. There is something it is like to be in a particular mental event with the property of intentionality, and there is the fact that it represents something. The former aspect is the phenomenology of that mental event, the latter is its intentional property. It should not be assumed that there is any intimate connection between these two properties. Of course, if it was thought that intentional properties are identical with, or constituted from, phenomenal properties, then the conclusion that we come to with regard to phenomenal properties in general will apply to them. If these types of properties are distinct then our approach is directly vindicated, for we shall discuss them as if they are distinct.

The second point concerns the extent to which our focus upon these two types of properties will be of relevance to the first challenge of epiphenomenalism that we decided to examine. Some might resist the focus on properties per se, preferring to see events in a different way to the one adopted herein.⁸ For them, to put it crudely, properties are just to be understood as ways of describing events and only the latter are the genuine relata of the causal relation. If such a position is defensible then the threat of epiphenomenalism, and our answer to it largely passes them by, the only question is

whether they should be allowed to get off the hook that easily.

Equally, there are no doubt other properties of our mental life which people might cite as candidates for being non-physical, and therefore of significance to the argument. However, we have not the space to consider them all. Therefore, we shall just consider the question of whether mental events are physical or not with respect to these two. Clearly, if we did establish that these properties were either non-existent or physical our conclusion vis a vis the physicality of mental events would be conditional upon results in other areas.

The conditional nature of the conclusion would be absent in this regard if we found either of these properties both existent and non-physical. In that case, we would have established that, by our definition, mental events are non-physical. However, bearing in mind our reformulation of premiss (1), this would not be enough to establish that mental events are causally epiphenomenal because premiss (2) needs recasting. It should now read:

(2)' All mental events are wholly non-physical.

It would then depend upon whether the mental events partly composed from these properties had some other physical properties, as to whether they were causally epiphenomenal or not. Nevertheless, the two properties we have focused upon were chosen quite deliberately. It is a commonly

held intuition that mental events are causally efficacious in virtue of their phenomenal and intentional properties. So, should we find the opposite, this would be a blow to our intuitive understanding of mind.

References

1. T. Nagel (1974), 'What is it like to be a bat?' Philosophical Review 83, (1979), 'Subjective and Objective' (both in his, Mortal Questions), (1986), The View from Nowhere (Introduction and Ch.1)
2. T. Nagel 'What is it like to be a bat' (Mortal Questions p.166)
3. T. Honderich (1988), A Theory of Determinism (p.77)
4. For this vein of accusation, see R. Rorty (1980), Philosophy and the Mirror of Nature (Chs. 1-2)
5. F. Brentano (1874), Psychology from an Empirical Standpoint (Book 2, Mental Phenomena in General, Ch.1, The Distinction between Mental and Physical Phenomena), (p.88).
6. T. Honderich (1988), A Theory of Determinism (Ch.2, p.83-86).
7. c.f. G. Frege (1892), 'On Sense and Meaning', (1918/19), 'Thoughts' (in his Collected Papers, edited by B. McGuinness).
8. e.g. D. Davidson (1970), 'Mental Events' (L. Foster and J.W. Swanson eds., Experience and Theory), (1967), 'Causal Relations' Journal of Philosophy 64

Chapter 5

On the Existence of Phenomenal Properties

One of the problems that faces any discussion of phenomenology is the controversy which attaches to the question of how we should describe its character. In Chapter 4, we defined phenomenology to be the appearance of the mind to the subject whose mind it is, remaining neutral over whether it was an entity composed of phenomenal properties. In the course of this chapter, we shall consider various arguments that have been put forward, as to the status of the appearance of mind. Our aim will be to establish that there exist some kinds of phenomenal property to which the argument of Chapter 6 may appeal. So, we should not be taken to be committed to their existence generally, although it is difficult to know where to draw the line.

(a) **The Basic types of Phenomenal Properties**
and an argument for their existence.

We may divide the phenomenal properties into the following categories. First, there are those which are attached to sensations, such as pain. Second, there are

those phenomenal properties associated with our moods, and emotions. Under the latter, we will also include the subjective feeling of certainty that attaches to a belief. A third type of phenomenal properties attaches to our occurrent cognitive events such as thoughts and beliefs. When we entertain a thought, or consider one of our beliefs in such a way that it seems legitimate to say that we are conscious of thinking or believing something, there are associated phenomenal qualities. Some of these qualities will attach to what we take to be the vehicles of thought such as sentences. A fourth type of phenomenal properties are those which attach to our imaginings and, more generally, the images which pass before the mind. The fifth and final type of phenomenal qualities are those attached to perceptual experiences of all sorts. It is here that the most controversy tends to occur. For, some would argue that the putative appearance of the mind in this case is just the appearance of the world.

The argument for taking each of these phenomenal properties as part of the appearance of mind, rather than of the external world, is as follows. In every case, it is natural to say we feel something. What we feel may either be an object in the world, such as the smoothness of another's face, or a constituent of our conscious mental life. It seems implausible for most of the cases indicated, that they are constituents of the world.

Therefore, if they exist, they are part of our mental lives and constitute, thereby, part of the appearance of the mind.

The argument for their existence is straightforward. It is that in enumerating these five types of phenomenal properties we are pointing to something which each person can recognise on the basis of their experience of their own mental lives. We all know what we are talking about to some minimal extent. This is a prima facie reason to take what we are talking about to exist. If we take ourselves to discriminate something, identify it, talk about it and, most basically of all, take ourselves to experience it, the putative thing which is the object of all this attention exists, unless there is good reason to believe otherwise.

Of course someone can deny that he or she knows of what is being spoken. Then there are only two alternatives that face that person: either they must hold themselves to be peculiar, or they must believe they have seen through the radical misconceptions of others. If the latter belief is held, they are obliged to provide an error theory, a theory, in particular, of why so many people have taken phenomenal properties to exist, when in fact they do not. If they adopt the former belief, and take themselves to be peculiar, their behaviour and activities drops out of this inquiry. We are interested

in whether phenomenal properties are epiphenomenal. Since they have no such properties, their mental life and behaviour is of no interest to the inquiry, whatever challenge they bring to our notion of mental events. The rest of the chapter will be concerned with the discussion of three types of error theory.

(b) **The Adverbial Approach to Phenomenal Properties**

The adverbial approach to phenomenal properties, if it is successful, establishes that there are no such properties. All five types of phenomenal properties, we identified, do not exist. Since the adverbial approach is most successful in its treatment of sensations, we shall consider it in this area. Should it be found wanting there, it will be found wanting everywhere. Should it be successful, we will get some idea of how the theory works, and how it may be extended.

Consider the sentence

(1) George feels a stabbing pain.

One way of understanding this sentence is by taking it to have the following logical structure.

(2) $SP(g)$

With the interpretation:

$SP(x)$: x feels a stabbing pain

g : George

However, if one represents the logical structure in this

way one loses the following arguments which seem intuitively valid.

(3) George feels a stabbing pain, therefore George feels something.

(4) George feels a stabbing pain, therefore George feels a pain.

To capture the logical form which makes (3) valid we have to provide something like the following:

(5) $F(g, sp)$

with the interpretation:

$F(x,y)$: x feels y

g : George

sp : an individual stabbing pain,
that named by 'sp'

For, from (5) we may then derive, by existential generalisation,

(6) $(\exists x) F(g,x)$

However, the suggested account of the logical form would not deal with the validity of the inference designated (4). For this, something of the following order is required.

(7) $F(g,p)$ and $S(p)$

with the interpretation

$F(x,y)$: x feels y.

g : George

p : an individual pain, that named by 'p'

$S(x)$: x is stabbing

The problem with both (5) and (7), as far as many are concerned, is that on the standard interpretation of quantification, the 'objectual' interpretation, we are committed to the existence of individual felt pains. At least, we are so long as the following is granted to be an acceptable inference.

(8) $Fa \rightarrow (\exists x) Fx$

where 'a' stands for an arbitrary name.

Those who are committed to the claim that everything is physical find themselves uncomfortable with the conclusion that there are individual felt pains, which we may reasonably think of as occurrences of certain phenomenal properties. They suppose that such 'individuals' are obviously non-physical. If this is 'obvious', then admitting their existence is admitting that not everything is physical.

From our own point of view, it is not true that the occurrences of phenomenal properties are obviously non-physical. So, the inference from the claim that occurrences of phenomenal properties exist to the claim that not everything is physical need not be possible. However, since we provide an argument for the non-physical nature of the occurrences of phenomenal properties in Chapter 6, an argument we may be partially convinced by, it seems reasonable for the proponent of the claim just

mentioned to be anxious.

The adverbial approach¹ is designed to show that the inferences just described above are valid, yet at the same time, that we need not be committed to the existence of the offending individuals. The basic idea is that every sentence of the same type as (1), one which contains a putative two-place predicate such as '- feels -' can be rephrased as the following:

(9) George senses F-ly

where the predicate '- senses F-ly' is taken to be primitively true of an individual. The assertion of a sentence containing just the predicate and the term referring to the individual is said to ontologically commit one only to the individual since there is no account to be given of what makes a predicate true of the individual. The search for such an account is viewed as misguided.

All descriptions of what one senses are taken to be operators on predicates. An operator upon a predicate takes a predicate upon which it operates and forms a more complex predicate in turn. The primitive predicate is '- senses'. This we may take to be true of an individual, in this case George, no matter what the individual is sensing. Suppose we wished to describe 'George feels in pain' in this form. We would write

(10) George senses painfully

Or, in logical form,

(11) [P]S (g)

with the interpretation

S(x): x senses

g : George

[P] : painfully

The adverbial approach is known as such because it makes all terms referring to types of sensations adverbial modifications of the verb 'senses'.

We can capture the inferences mentioned at the beginning of the discussion by the introduction of special rules of inference concerning operators. Suppose that O_i is a schematic letter in place of which are substituted letters standing for operators. Then we may say that the following is an acceptable rule of inference:

From any logical schema of the form, $[O_1] S(x)$ we may draw the inference $S(x)$

Equally, the following:

From any logical schema of the form $[O_1] [O_2] S(x)$ we may draw the inference $[O_2] S(x)$.

Both are known as operator detachment rules.

We would represent (1) as

(12) [St][P] S(g)

with the interpretation

[St] : stabbingly

[P] : painfully

S(x) : x senses

g : George

It is clear that the rules of inference specified above would allow us to make the second of the required inferences. Of course, there are complexities of inference which the account we have sketched, drawn from Michael Tye's work,² does not capture.³ However, the complexities may be captured by further strategic introductions of operators and inference rules.⁴ Of this, there is little doubt.

The inference recorded in (3) may only be captured without an ensuing ontological commitment if the existential quantifier in this case can be interpreted substitutionally. Crudely speaking, if we interpret the existential quantifier in this way ' $(\exists x)Fx$ ' should be read, there is at least one name 'a' we can substitute for the variable such that 'Fa' is true. There is no commitment to an object named. On the other hand, it might be denied that we can draw the inference. Some philosophers⁵ have dubbed the putative inference an instance of the 'phenomenological fallacy'. They hold that it is illegitimate to suppose that our introspective reports of the appearance of our mind are descriptions of the occurrence of phenomenal properties. Given the additional assumption that the existential quantifier may only be interpreted objectually, we get the conclusion

that the legitimacy of the inference outlined in (3) should be rejected.

We have seen that there is a way of rewriting sentences and introducing inference rules that enable us to avoid ontological commitment to phenomenal properties. We have also seen that ontological commitment may be avoided by adopting a substitutional interpretation of quantifiers. However, it is questionable whether the existence of these techniques helps us with the question of what exists. Their proponents consider that the techniques have limited application, but rarely reveal what linguistic facts make them especially applicable to talk of phenomenal properties rather than, say, talk of spatial properties. If there were no reason to believe in the existence of phenomenal properties other than the patterns of inference we have been describing, perhaps it would follow that the successful casting of these patterns into language that did not come with the purportedly unwelcome ontological commitments would throw the existence of these properties into doubt. However, in the first section of this chapter we provided other reasons for believing in the existence of phenomenal properties, therefore this does not seem to be the correct way to view the situation.

We should also recognise that the ability we have to recast inferences in ways that avoid ontological

commitment may well extend further than it should, if it were taken to be a sensitive measure of what in fact exists. The existence of unnamed entities has often been cited as an objection to the uniform adoption of the substitutional interpretation of quantifiers. We want to take 'Something is F' as true, even if nothing with the property F has been named. Our ontological commitments are thought to follow from this fact. But, it would seem that a language developed by God would not lack the relevant names so could have its quantifiers interpreted in that fashion and yet one would not wish to believe that God thereby had no reason to believe in the existence of entities. Of course, God will have other reasons to believe in their existence, but then, as we have already noted, so do we. The important point, however, is this: if one's linguistic ability will in part determine whether the working of one's language gives one reason to believe in the existence of certain entities, it seems probable that our ability to paraphrase our language into non-ontologically committing forms will not match up with what in fact exists.

One reason that has been offered in favour of the adverbial approach is that to allow the existence of phenomenal properties is to recognise that a whole collection of awkward questions are legitimate.⁶ As far as one can assess such a claim it seems that by parity of

reasoning we should adopt an adverbial approach to philosophy. Surely an approach cannot have as its sole recommendation the fact that it allows one to avoid awkward questions unless some reason is given for thinking that the questions are not legitimate in the first place.

In his attempt to undermine an argument like the one advanced in the first section, Tye says that introspection only provides one with the materials for concluding that 'I am in pain' is true or false. It does not enable us to determine whether there really is a felt quality of pain. In which case, Tye has to give an account of what grounds a person has, from introspection, to assert that it is true that he or she is in pain, rather than thirsty. Unless there was good reason to think otherwise, it seems overwhelmingly tempting to suppose that the distinct judgement one makes that one is in pain is based upon one's introspective ability to detect the existence of the feeling of pain. Now, it should not be assumed that one has a right to assert that just such introspective abilities are possessed, one does not. Nevertheless, we can urge that unless some reason is given for doubting this interpretation of introspection, it provides some grounds for believing in the existence of phenomenal properties.

(c) Does an allegiance to some combination of empiricism, realism and pragmatism throw into doubt the existence of phenomenal properties?

A crude characterisation of empiricism is that it is the philosophical doctrine which holds:

An essential source of justification for asserting or denying a certain type of proposition is sense experience.

Upon this view, empiricism concerns the justification of our beliefs and not our possession of concepts. To say that sense experience is an essential source of justification is to say that whatever other justification is cited one necessary component of our justification in believing a proposition is sense experience. This is because sense experience is taken to reveal to us the world the type of proposition in question describes.

An equally crude characterisation of realism would be that:

The existence and nature of entities is logically independent of our knowledge of those entities. Thus, it is possible that some proposition concerning these entities is true, yet nobody knows that it is true.

Realism is a thesis that may be adopted for some entities, and not for others, as is well known.⁷ It allows for the possibility that we are fallible. However, a straightforward belief that we might be fallible is all

that is strictly needed for the argument in this section to proceed.

The combination of the adoption of empiricism, and realism with regard to a certain range of entities yields the conclusion that we may be mistaken in those propositions we assert about the entities in question on the basis of sense experience. Thus, if the two theses are adopted with respect to phenomenal properties, it follows that the claims we make on the basis of our experience of phenomenal properties may be mistaken. Notice that some sort of belief in the possibility of our fallibility, is crucial for the conclusion just made. If one were just an empiricist, one might hold that experience of phenomenal qualities gave us an infallible insight into the nature of phenomenal properties.

So far all we have concluded is that it is possible we may be mistaken about the nature and existence of phenomenal properties. From that it does not follow that we are so mistaken. However, the following line of argument has seemed tempting to some.⁸ If we are wedded to a notion of phenomenal properties which are not susceptible to investigation by science we would be justified in considering whether we should abandon the propositions to which we are committed in possessing the notion. Past progress in science is taken to be indicative of the potential for future progress, and past

misconceptions which, for a while held up the progress pointed to, are taken as a model for the likely status of our current conception of phenomenal properties. Since this conception throws into doubt the future progress of science, which we would otherwise have reason to expect, we should take the problem phenomenal properties pose, admitted by nearly everybody, as reason to believe that they do not exist, or, at least, that they do not exist as conceived.

The line of argument just sketched is stronger if combined with a form of pragmatism. Let us take it that the pragmatist holds:

The only justification one may have for believing a proposition is that in so doing we find the goals that we have adopted advanced.

A scientific pragmatism would be one which allows that we should alter our beliefs in accordance with scientific goals. It is reasonably clear how this thought would serve to bolster the line of argument sketched above.

The philosophical doctrines of Realism, Empiricism and Pragmatism each have a significant body of argument and rhetoric to recommend them. Therefore, if it were the case that inspection of scientific progress plus their adoption yielded the conclusion that we are mistaken about the existence or nature of phenomenal properties, then there would be some cause for concern. How much concern

would, of course, depend upon a more detailed examination of scientific progress, and a more precise formulation of the arguments and doctrines to which we have adverted. However, there seems to be a general argument that renders, at least, questionable the whole approach.

Something which characterises both empiricism and, arguably, scientific pragmatism is their emphasis upon the evidence one can obtain through experience. If certain evidence can be obtained in most cases where one would expect that it should obtain, as a result of theories one has already adopted, or induction from past experience, then the evidence may be called stable. Stable evidence has the most status. Our experience of phenomenal properties, if it is not denied to be experience of these putative things, is stable evidence that these things exist and have a certain nature. Therefore, by the lights of empiricism and scientific pragmatism one should believe in the existence of phenomenal properties and the nature one apprehends them to have.

The argument of the preceding paragraph suggests that one cannot use the antecedent plausibility of empiricism and scientific pragmatism to throw into doubt the existence of phenomenal properties and the nature we take phenomenal properties to have. Whatever force these doctrines provide when taken in conjunction with some rather nebulous remarks about scientific progress, is

entirely undermined when one considers other commitments of these two philosophical doctrines. Unsurprisingly, those who put forward the line just described question the claim that was made in order to throw the approach into doubt. They wonder about the legitimacy of our experience of phenomenal properties. The latter part of this section and the next will be devoted to assessing arguments that have been offered to undermine our faith in such experience. It is crucial to realise that their whole approach depends upon these arguments.

Before we can consider the first of the arguments, we need to note a difference between our approach and those of the philosophers who put forward the argument. We have taken phenomenal properties to be what constitutes the appearance of the mind. We have made no claims about their nature apart from this. We have not, for instance, held that they are infallibly known, or that they have peculiar properties physical properties cannot have. All of this, so to speak, is up for grabs. So, when a philosopher, such as Richard Rorty⁹, says that reports concerning phenomenal properties, are just reports concerning the neural properties of brain events and not reports concerning genuinely phenomenal properties, as far as we are concerned there are two separate claims being made here. One is that reports of phenomenal properties are reports concerning the neural properties of brain

states. There is nothing we have said so far, which brings us into disagreement with this claim. The second claim is that the reports which are made are not made of genuinely phenomenal properties. It is just this that we are seeking to reject and it is the arguments for this claim in which we are interested.

Now, it may be said that because we have adopted this minimal understanding of phenomenal properties our position is not in contention with Rorty's. Certainly, this should be said with regard to Paul and Patricia Churchland's approach. It is because Rorty often says that he is determined to reject the view that appearance is constituted from properties that it seems reasonable to still view our position as in opposition.

The argument that Rorty uses, to bolster his position, and that Paul Churchland¹⁰ also adopts, but only to oppose a particular view of phenomenal properties, is drawn from Wilfred Sellars' work.¹¹ The argument focuses upon the relationship between language and the world. Suppose for the sake of argument that the concepts that we possess are determined to a large extent by the language we speak. If it could also be shown that the language we speak determines what we take ourselves to be aware of, in the world, then the following possibility seems to open up. If we change the way we talk of things, we will gradually become aware of different types of things. So,

the idea is that we only take ourselves to be experiencing phenomenal properties of a particular character because we talk that way. Give up that way of talking and we will soon see that what seemed a stumbling block to neuroscience was just inveterate chatter. What can be the rationale for such a view?

The 'myth of the given', according to Wilfred Sellars, is the thesis that we can make infallible judgements concerning the nature of our present sensations. His argument against this 'myth' is that: first, our apprehension of, in our terminology, the phenomenal property associated with a sensation is not an apprehension of a universal but rather is of the occurrence of a property; second, that what enables us to go beyond this apprehension of a particular is our linguistic abilities. Here, the notion of a universal is taken to be of something for which it is logically possible to occur at more than one spatial location at any one time. The alternative suggestion Sellars is making is that it is our capacity to use a language which enables us to talk as if there are resemblances and recurrences in nature. If Sellars is right in making these two claims, then it may look as if we have some support for the position we oppose. It is not even essential that Sellars be right about the connection between language and concepts, so long as the second claim can be reformulated

in terms of concepts. Also, inessential to the line of argument we are considering is Sellars' Nominalism, the view that there are no universals. Instead, we can take Sellars' point as entirely epistemological. Even if there were universals, we can say, there is still the problem of whether we can apprehend them as such. The fundamental claim, then, is that we can only apprehend kinds of things as of those kinds, for instance, as phenomenal properties, by possessing certain linguistic cum conceptual abilities.

Suppose we accept for the purposes of argument that we think we grasp what we take to be universals as a result of language. The obvious question to ask is how does language alone enable us to group things. When one looks at the mechanics of the account it may look as if it cannot, without ascribing to us the ability to grasp at least some universals independent of language, those of predicates. For example, consider the application of the predicate '- is square'. According to the hypothesis we need to ascribe no ability to apprehend the universal of squareness, and, indeed, could not. All we need to do is note that for each square thing the predicate '- is a square' applies to it. It is just our ability to use this predicate which enables us to group them accordingly. But, to be able to group all square things together you must be able to group all applications of '- is square' together, and how can you do that without suggesting that

the subject has the ability to apprehend some universals, namely, predicate types? If the same denial that we can apprehend universals is made here, the issue just arises once more with the predicate we use to group together these predicates, the predicate "- is the predicate '- is square'".

One reply that Sellars could make at this point is that some universals are more difficult to apprehend than others. However this would be a fairly desperate move. Instead, Sellars¹² would probably say that although we take ourselves to be grasping universals, in the form of types of predicates, quite independently of our language use we in fact are not. We are, once more, just under the illusion that we are. Our abilities to group things are always to be understood as abilities to apply predicates in the way we have described. However, unless we suppose that we have an infinite hierarchy of abilities to apply predicates to predicates, it seems that ultimately either we must allow we do not in fact have the ability to group things or that we apprehend universals after all. It is at this point that we can see the utility of Sellars' espousal of a radical nominalism. By holding that there are no universals to be grasped he effectively rules out the possibility of our succumbing to temptation. Now, the debate concerning nominalism tends to centre around what may be taken to be genuinely explanatory of our experience

and abilities and what may not.¹³ However, we need not enter into such a debate. It is not by adopting nominalism that the existence of phenomenal properties is threatened. Nominalism treats all properties alike and allows much the same issues to arise concerning the types of properties there are, albeit in different terms.

The argument above is meant to undermine the thought that, in any sense, we perceive the world as it is, rather than just in the way we conceive it to be. However, if this genuinely is the position, it is hard to see how the proponents of it could maintain their adherence to some form of empiricism for now the revelations of sense experience are being down peddled in favour of our conversational preferences. So, to what extent are the philosophers we have mentioned empiricists? Matters are complicated by the fact that each professes not to be an empiricist, but takes this doctrine to be that one's immediate reports concerning sense experience have some sort of infallibility, and enjoy a justification independent of other beliefs which do not concern sense experience. We can, for the sake of argument, agree with them that such a form of empiricism is indefensible. However, this still leaves us with the question of whether they are empiricists in our sense, a sense we claimed was in tension with their doctrines.

It may seem that given that this tension exists, our

answer is obvious. Unfortunately, it is not. Their emphasis on the scientific approach, and the fact that certain theories work better, and explain more things, suggests that they do place some weight upon sense experience. How may one determine whether a theory works better unless there is something in experience to which one appeals? There is at least a prima facie case for supposing that they are empiricists in the weak sense we have specified. Moreover, it is not just that the adequacy of theories is in part determined by the character of sense experience, also, the problems for which the theories are solutions must be derived from experience.

Some of the philosophers we have discussed, in particular, Paul Feyerabend and Rorty, in his later work, seem either tacitly or explicitly to recognise these points and are inclined to take science as a less rational and more ideological activity, giving up any allegiance to empiricism. To a position of this sort, we have no answer because we could have no answer. There are no 'reasons' that we may offer which are going to be accepted.

However, the aim of our inquiry was to assess the likely role of mental events in psychological explanation, given that their role was coming under threat from certain empirical evidence, or from certain lines of reasoning. Within this framework, it seems we do have a reply to

those who question the existence of phenomenal properties. It is that whatever weight is given to the information we obtain from experience, the stability of our experience of phenomenal properties would command a degree of scientific attention that should outweigh the degree of support there is for our scientific theories. So, as yet, we have been provided with no reason for thinking that, in particular, our experience of phenomenal properties is to be questioned.¹⁴

(d) **Wittgenstein on Phenomenal Properties**

The most famous argument against the existence of phenomenal properties derives from Ludwig Wittgenstein.¹⁵ However, it is more precisely seen as an argument against the view that phenomenal properties are intrinsic properties. The proponent of this view of phenomenal properties will hold that their occurrence is metaphysically independent of the occurrence of anything else. A consequence of this view is that the occurrence of phenomenal properties is metaphysically independent of what kind of judgements a subject is disposed to make as a result of them. It is to this consequence that Wittgenstein's argument is addressed.

Suppose a subject was so constituted that the phenomenal properties that were instantiated as part of his or her experience, when, say, he or she was faced with

a red object, constantly changed. However, the subject did not notice this because his or her memory compensated for this continual alteration. If I were the subject then I would not notice that anything is happening to me. Faced with red objects, one phenomenal property would be instantiated as part of my experience and then another, but I would not judge that the object has changed colour because my memory will be that the object had the colour I am now perceiving it to be.

For those who believe that phenomenal properties are intrinsic properties, the subject we have just described could exist. The example appeals to this notion of phenomenal properties. However, we are supposed to conclude that the case envisaged is not possible. It is supposed to be a *reductio ad absurdum* of the account of phenomenal properties we are considering. If phenomenal properties can only be intrinsic properties, then the success of the argument would not only demonstrate that phenomenal properties were not intrinsic properties but also that phenomenal properties did not exist.

How is the argument supposed to work? Presumably as follows:

- (1) If phenomenal properties are intrinsic properties, then the case described is possible.
- (2) The case described entails a contradiction.
- (3) What entails a contradiction is itself not

possible.

(4) The case described is not possible.

Therefore:

(5) Phenomenal properties are not intrinsic properties.

The crucial step in the argument is clearly contained in premiss (2).

The contradiction is, at first glance, a little hard to identify. There is nothing in the case as we described it above that seemed contradictory. However, the contradiction seems to arise from the following two claims concerning the identity conditions of phenomenal properties.

(a) Two phenomenal properties are identical if and only if they feel the same to the subject in whose experiences they occur.

(b) The identity of phenomenal properties is in no way determined by the kind of judgement the subject is disposed to make on the basis of their occurrence.

The case we described is an articulation of the second point, but putatively runs directly counter to what is recorded in (a). The case dramatises the contradiction that lurks in the joint holding of these two theses. All that one has to recognise, for the contradiction to be manifest, is that two phenomenal properties feel the same

if and only if they tend to give rise to the same kind of judgements.

Now, there are various points that one can make about the legitimacy of (a). There are also things to be said about the link between 'feeling the same' and judgement. In addition, accounts may be offered of the nature of phenomenal properties which combine a limited intrinsic character, with some specification of their identity conditions in terms of judgement.¹⁶ However, we are not going to examine any of these replies to the argument just stated. All we shall do is question whether it is legitimate to suppose that if phenomenal properties exist, they must be intrinsic.

The most commonly offered reason for supposing phenomenal properties are intrinsic is that they appear intrinsic. But, Wittgenstein's example shows how phenomenal properties may not be intrinsic even if they appear that way. Therefore, we cannot appeal to our experience of phenomenal properties to demonstrate that they are intrinsic. Nor is there anything else to which we may appeal. However, if there is no way of establishing that if phenomenal properties exist then they are intrinsic, then it follows that showing that it is impossible that phenomenal properties are intrinsic does not entail that phenomenal properties do not exist. As there is no need to suppose phenomenal properties are

intrinsic, there is no threat to the existence of phenomenal properties from considerations that have their origin in Wittgenstein's writings.

In the course of this chapter, we considered the most often canvassed arguments against the existence of phenomenal properties. All the arguments we examined were unsuccessful, in so far as they were relevant to our inquiry. Therefore, within its scope we have found that there is some reason to believe in the existence of phenomenal qualities and no good reason to disbelieve in their existence. Therefore, it would, overall, be reasonable to believe in their existence.

References

1. Famous proponents of which are: R. Chisholm (1957), Perceiving (Ch.8); C.J. Ducasse (1951), Nature, Mind and Death (Ch.13); W.S. Sellars (1966), 'Reply to Aune' (H. Casteneda, ed., Intentionality Minds and Perception), (1975), 'The Adverbial Theory of the Objects of Sensation', Metaphilosophy 6, (1968), Science and Metaphysics (Ch.1).
2. M. Tye (1989), The Metaphysics of Mind (especially Ch.3).
3. See Frank Jackson (1977), Perception (Ch.3) for some examples.
4. See M. Tye (1989), (Chapters 4-5).
5. e.g. U.T. Place (1956), 'Is Consciousness a Brain Process?' British Journal of Psychology 47.
6. M. Tye (1989), (Ch.3).

7. c.f. M.A.E. Dummett (1982), 'Realism' Synthese 52.
8. P.M. Churchland (1979), Scientific Realism and the Plasticity of Mind (Ch.1-4), (1985), 'Reduction, Qualia and the Direct Introspection of Brain States', Journal of Philosophy 82, (1989), 'Knowing Qualia: A Reply to Jackson' (in his Neurocomputational Perspective), (1986) 'Some Reductive Strategies in Cognitive Neurobiology' Mind 95; P.S. Churchland (1986), Neurophilosophy (Chs. 6-8); P.K. Feyerabend (1963), 'Materialism and the Mind-Body Problem' The Review of Metaphysics Vol. 17, (1963), 'Mental Events and the Brain' Journal of Philosophy 60; R. Rorty (1965), 'Mind-Body Identity, Privacy and Categories' and 'In Defence of Eliminative Materialism' both in The Review of Metaphysics 19, (1980), Philosophy and the Mirror of Nature (Ch.1-2).
9. See previous references to Rorty's work in footnote 8.
10. P. Churchland, see references in footnote 8.
11. c.f. W. Sellars (1956), 'Empiricism and the Philosophy of Mind' (in H. Feigl and M. Scriven, eds., Minnesota Studies in the Philosophy of Science Vol.1, and reprinted in his (1963), Science, Perception and Reality).
12. c.f. W. Sellars (1969), 'Language as Thought and as Communication' Philosophy and Phenomenological Research 29.
13. c.f. D. Lewis (1983), 'New Work for a Theory of Universals' Australian Journal of Philosophy 61.
14. For further discussion of this matter, and, indeed, many of the issues in this chapter, see H. Robinson (1982) Matter and Sense (Ch.5).
15. L. Wittgenstein (1953, 1958), Philosophical Investigations (especially I s.293 and p.207).
16. S. Shoemaker (1975), 'Phenomenal Similarity' Critica 7.

Chapter 6

An argument for the proposition that phenomenal properties are not physical

There are many arguments that have been offered for the claim that phenomenal properties are not physical. During this chapter, we shall first consider, in general terms, what one would have to demonstrate about mental properties, to establish that mental properties are non-physical. In coming to understand what needs to be established, we will appreciate what is required to demonstrate that phenomenal properties, in particular, are non-physical. It will be claimed that once one recognises what is required, the arguments hitherto offered for this conclusion are problematic. In their place, an argument will be developed which attempts to avoid the pitfall to which other famous arguments are susceptible.

(a) **What would we have to show, to show that mental properties are non-physical?**

Suppose that some account of the nature of the physical can be provided. This, in turn, will provide us with an understanding of the nature of the non-physical, for, with one exception, by definition, what is not

physical will be non-physical. In order to find out how we should show that a particular entity was non-physical, it would be better if we first examined how we should show that it was physical. Specifically, we shall consider the case of mental properties. In so doing, we will bring to light a problem which plagues both those who believe that mental properties are physical, and those that do not, namely, that contrary to the impression that is sometimes given, it is very difficult to provide grounds for believing either of these claims.

The obvious approach to the problem would be to demonstrate that the principles by which we categorise the physical as physical, yield the conclusion that mental properties are also physical. For instance, if we wished to show that a certain object was a cube, the measurements, made in other cases to establish this fact, would be engaged in once more with respect to the new object. If the same relations hold, the object is a cube. A similar procedure is envisaged for discerning whether something is physical.

Unfortunately, there are complications. One arises if it is thought that a necessary condition upon what is physical is that it has a single spatio-temporal location. Such a view may seem to come unstuck when we turn to properties. Suppose one believes that properties are 'abstract' and thereby not spatio-temporally located, or,

at least, do not have single spatio-temporal locations, in contrast to their occurrences. Yet it might be thought that properties are still physical. After all, is not the property of mass a physical property?

The immediate reply is that we must distinguish between two ways in which properties may be physical. First, there is the matter of whether they as properties are physical. If properties are abstract then it follows they will not be physical given the correctness of the claim about the physical mentioned above. Even the property of mass is not physical from this perspective. The second way in which properties may be physical is if all their occurrences are physical. It is in this second sense that the property of mass is uncontroversially physical, for it is claimed that whenever such a property occurs it satisfies the intuitive criterion that determines whether or not something is physical.

Unfortunately, the complication we have just identified is not the only one we face. A second concerns the existence of what we may call 'neutral' properties. Let us stipulate that unlike physical and non-physical properties, neutral properties may have both physical and non-physical occurrences. Any case is going to be a controversial illustration of the thesis that there may be neutral properties, but the temporal relations expressed by the predicates '- is before -' and '- is after -' may

be examples. Suppose one believed that mental properties were non-physical, and that their non-physical occurrences had temporal relations to each other. Suppose one also believed, as is usual, that there are temporal relations between occurrences of physical properties. Then one might say that since temporal relations have both physical and non-physical occurrences depending upon the nature of the relata, the relations themselves are neutral. The coherence of this description suggests that there is room to allow for this way of describing things.

The possible existence of neutral properties encourages us to distinguish between attempting to show that mental properties are physical, and attempting to show that at least some of the occurrences of mental properties are physical. It is not important for us to determine whether or not mental properties are neutral properties. We may proceed in a way which does not prejudge this issue, yet which is satisfactory from the point of view of our inquiry. The argument we are considering for some form of epiphenomenalism does not require that mental properties be physical, just that the occurrences of mental properties in us are physical. Therefore, if we can show that mental properties can have some physical occurrences, then most of the battle would be won. Given that we have a prior conviction that mental properties as they occur in us are not causally

epiphenomenal, then we would have reason to believe that these properties have physical occurrences in our case. Thus, we would be entitled to reject premiss (2)' of the argument. Mental events are not wholly non-physical.

A third complication concerns the distinction sometimes made between showing that occurrences of mental properties are physical, and showing that mental predicates have physical conditions of application.¹ A mental predicate is here defined to be any which we intuitively take to ascribe a mental event, state or property to a subject. If one demonstrates that occurrences of mental properties are physical, it follows that mental predicates have physical conditions of application. However, the reverse entailment does not hold. The reason for this is that the concepts expressed by mental predicates, 'mental concepts', may be concepts of entities that do not possess the features that physical entities possess or are not concepts of entities at all. It may, nevertheless, be the case that the concepts in question do apply in certain given physical circumstances. If mental concepts are concepts of entities, then their use may rest upon a mistake about the way the world is. Still, by having a systematic application, the concepts may have a utility. So, it need not follow that we should abandon our use of mental predicates in these circumstances. However, we might have to revise them to

the extent that we take them no longer to be concepts of entities.

If mental concepts are not concepts of entities then the mental predicates which express them have rules of application that make no reference to the type of features an entity should possess for them to be applicable. Or, at least, what reference there does occur is not viewed as the specification of a necessary condition as to when the predicate applies. It may be a sufficient condition. The concepts expressed by the predicates in question would, typically, be viewed as a way of conceiving the world only in terms of the concept user's interests, or intellectual proclivities, rather than also responsive to 'objective' features of the world.

We shall not presume that the distinction just adverted to is legitimate, however it would be unwise, in our present circumstances, to presume the opposite either. What we should recognise is that if mental concepts have physical conditions of application, then once more the argument outlined at the beginning of Chapter 1 will be unsound. It will not be true that mental things are non-physical. So, even if it cannot be shown that occurrences of mental properties are physical, it still might be possible to avoid the conclusion of the argument. This qualification must be born in mind in subsequent discussion.

Our reformulated question is thus as follows, 'How would we show that some occurrences of mental properties are physical occurrences or that mental predicates have physical conditions of application?' In the case of mental properties, we would have to provide an accurate and exhaustive description of their nature that would then be compared with our account of the physical. The description would have to be both accurate and exhaustive otherwise there would always be the danger that we had passed over that part of the property that indicated either that it could or that it could not have physical occurrences. In the case of determining whether mental predicates had physical conditions of application, a similar sort of description would be required, but there we would be attempting to articulate the concepts we expressed by such predicates, and the aim would be to show that the concepts had a character such that it was either possible or not possible that they should have physical conditions of application.

If we could be certain that the descriptions of which we have spoken are only to be used to articulate the concepts expressed by mental predicates, then it may be legitimate to suppose that such a description will be available to us a priori. This would, of course, depend upon how one understood the notion of the a priori, and its relationship to our mastery of concepts. However,

since we, also, wish to determine whether mental properties have physical occurrences, we should not suppose that a description of the requisite sort will offer itself to us a priori. Moreover, we could have misconceptions about the nature of mental properties.²

So, we should consider quite carefully how we are going to draw up the type of description which we said was necessary. It might be thought that to attempt to provide a description of mental properties is to presume that mental properties are complex. A description of something may be thought to identify features of that thing, and thus, simple properties could only be 'described' by general terms referring to their nature in its entirety. They have no aspects which may be singled out.

Put like this, the potential objection cannot be quite right, even granting it some intuitive force. It should not be assumed that what we conceive of as simple is simple, and vice versa. Thus, for instance, it may be possible to provide a description of something in terms of aspects of our understanding of that thing even though the thing is simple. Our concepts would unite to at least partially correctly describe a simple thing, so to speak, even though we would not conceive of it as simple.

However, the main problem with the objection mentioned lies elsewhere. Descriptions are available even if a property is simple. A description may refer to the

relations this property has to other properties. A statement of these various relations may then be used to single out one property rather than another. The suggestion is pretty familiar, and it is generally recognised that if the relations in question state the type of causal relations occurrences of the properties stand in, then the character of the properties will be to some extent illuminated. So the request for a description does not presume that the object to be described is not to be conceived of as simple.

What it would not be legitimate to suppose, without further argument, is that a statement of the causal relations which hold will tell us all there is to know about simple properties. To think it does is to move from the thought that the only description of a simple nature that can be provided is one general term referring to the nature in its entirety, and a description of the relations the thing stands in, to the thought that all that there is to be said can be captured by mention of the relations. It is just not clear why this move is legitimate.

The real problem that faces our approach is that it is difficult to see how one separates features that one takes mental properties to have as a result of our experience of mental properties, from features that they actually have. For instance, any argument that rests upon the intuition that mental properties cannot be neural

properties because the features we attribute to each are very different, or that claims that the features that the first has can be possessed independently of those that the second has, must at sometime face the objection that what is being asserted rests upon peculiarities of our access to properties of the respective types and not upon actual differences in the properties.³ Our introspective access is obviously quite different from any access we have mediated by sight, touch, and instruments which are the extension of these and other senses. It should come as no surprise that what we take ourselves to sense by these very different means, are very different. However, until we can filter out the contribution of these distinctive means of access, we cannot with any confidence assert that what we perceive as a neural property is not the self-same thing as that which we take to be a phenomenal property. The felt quality of pain may just look grey and feel spongy if one uses one's eyes and hands to detect it.

The difficulty here sketched reveals the current weakness of the position of those who claim that phenomenal properties are not physical. The weakness is further highlighted by the other qualifications we made. We noted that it would have to be shown that mental properties could not have physical occurrences nor mental predicates physical conditions of application. If mental properties were neutral properties we might expect that

our experience of physical occurrences of them would be subtly different from our experience of physical occurrences of physical properties. Yet, this difference would not imply that the properties in question were non-physical. A similar point can be made with regard to our conception of the physical conditions of application of mental predicates.

So those who wish to establish that mental properties are non-physical should not hope to demonstrate that our apprehension of mental properties reveals that these mental properties do not have physical occurrences, nor mental predicates, physical conditions of application. This would be, to put it mildly, an uphill task. Nevertheless, these remarks should not promote complacency on the part of the proponent of the claim that phenomenal properties are physical, have physical occurrences or are ascribed by predicates with physical conditions of application. A consequence of the difficulties we have mentioned is that their proposal is almost entirely unconstrained by the existence of conditions which, should they hold, indicate to us that the position adopted is false. This is intellectually unsatisfactory and renders the evidence for their position weak as a result. For these reasons it seems that a new approach is necessary. This will be the subject of the next section.

(b) The argument from awareness

The argument contained in the present section tries to show how certain features of awareness entail that phenomenal properties are non-physical. It explicitly attempts to avoid appeal to our experience of the character of phenomenal properties. At its crudest, it runs as follows.

- (1) An entity is physical if and only if
 - (a) It is spatio-temporally located, or composed from entities that are;
 - and,
 - (b) Its existence does not imply that a subject is aware of it.
- (2) The occurrences of phenomenal properties do imply that a subject is aware of them.

Therefore

- (3) The occurrences of phenomenal properties are non-physical

We have already argued in Chapter 1, for the truth of the first premiss. The focus will now be on the second. After we have presented the argument for the second premiss we shall consider, in the next section, various objections that may be brought to it. Such a discussion will be of particular importance because of slipperiness of the subject matter of awareness. It would be entirely reasonable to expect that the reasoning given involves

some kind of trick, or fundamental misconception. By discussing the various objections, we may hope to make manifest for the potential critic, exactly where the slip is being made, even if it is missed by us.

Why should we suppose that the occurrences of phenomenal properties imply that a subject is aware of them? What we shall try to argue is that certain facts of awareness can only be accounted for if we do suppose that phenomenal properties have this implication.

The first claim that we need to defend is the following:

- (1) If one is aware of something, then one must be partially aware of the nature of the state of awareness.

Suppose I am aware of the stapler that sits upon my desk. The claim applied to this case is that in order to be aware of the stapler I must be partially aware of the nature of my awareness. The notion of awareness to which we are appealing here is pretty weak. What holds for it will hold for all others. For a subject to be aware in this weak sense, he or she need not have conceptualised what he or she is aware of, or be inclined to make judgements. All that is necessary is that the subject may make primitive responses to the existence of the object on the basis of his or her sensory awareness. For example, by avoiding the stapler if it is thrown at him or her. It

is the type of minimal awareness most people illustrate by talking of how they sometimes drive a car.

It would be helpful at this point if we introduced a piece of terminology. Let us call the object of which a subject is aware, the 'object of awareness'. We shall then call that part of the subject that is his or her awareness of the object, the 'state of awareness'. The phrase is meant to, in neutral terms, capture the fact that the subject in question is aware of something. It is not intended to be illuminating but just to pick out some component of the subject, that we would like to know very much more about, by talking of what we suppose it involves.

Why should it be thought that an awareness of other things must partially involve an awareness of the state of awareness? Well, the thought is that if one were not aware of one's state of awareness then one could not be aware of the object of the state of awareness. The state of awareness would have occurred, so to speak, without the subject, whose state it is, being aware, in our minimal sense, that it occurred. But if the subject is unaware of the state of awareness how could the state be an awareness, on the part of the subject, of the object already mentioned. It seems it could not. Thus, we have the required conclusion that if one is aware of something then we must be partially aware of the nature of the state

of awareness.

It is at this point that we have to examine our notion of a 'state of awareness' a little more carefully. So far, we have just taken it to report a fact about a subject. However, it may be thought to have 'explanatory' force. Upon this view, when we ask for an explanation of why somebody is aware of something, we obtain as an answer, that he or she is in a state, that of awareness, such that it is because he or she is in such a state, that he or she is aware of something. With this understanding of the role of a state of awareness in human psychology, at least, it seems we should understand the first claim that we are defending as the assertion that in order to be aware of something, we must be partially aware of the nature of the state of awareness, otherwise we would have no explanation of how we were aware of something in the first place.

However, to adopt this approach is to leave oneself open to a difficulty noticed by Gilbert Ryle,⁴ albeit in a slightly different context. Ryle's thought transposed was this: if in order to be aware of an object we must be aware of the awareness then it looks as if to be aware of the awareness one must have another awareness on top of that, and so on. An infinite regress ensues. Moreover, the infinite regress is unacceptable in the present context because it defeats the motivation for introducing

the idea of a state of awareness in the first place. The motivation just mentioned was that we would in some way, no matter how minimally, explain how a subject was aware of something by saying that it was the object of a state of awareness. Now, we see that no such suggestion will do because we still would not have accounted for how a subject is aware in even this minimal sense. The same problem is raised once more, this time with regard to the state.⁵

The natural response at this point is to reject the tacit differentiation between our awareness of an object, the stapler say, and our awareness of the state of awareness. Instead, we might say that the state of awareness of the stapler doubles as a state of awareness of awareness and so on. After all, we may argue, it would not be a state of awareness if such questions as the one we just canvassed could arise. So in every case of awareness, there will be many states in one.

In principle, there is nothing wrong with this response, but it is important to recognise its limitations. We are told that one state of awareness, in effect, is many such states. In fact, it is an infinite number of such states, if the line of reasoning we have ascribed to Ryle is correct. But, the question still arises, what must the character of this state of awareness be, such that it solves all the problems concerning

awareness in one fell swoop? Or, to put the point another way, how can it be that there is such a state? The occurrence of words like 'must' and 'can' in the questions above should give you warning that you are going to receive a bit of philosophical encroachment on science, rather than a rigorous empirically established scientific theory.

It will come as no surprise that the answer offered here draws upon the distinction between those things whose existence implies a subject's awareness of them, and those things whose existence does not. For all those things whose existence does not imply that a subject is aware of them, there must be some extra fact which explains when a subject is aware of them, namely, the state of awareness. On the other hand, for those things whose existence implies that a subject is aware of them, the question does not seem to arise. Their existence and the subject's awareness of them go hand in hand.

The issue for us now is whether the connection between something whose existence implies awareness and a state of awareness is sufficient to resolve the problem identified by Ryle. Suppose we say that a state of awareness of some object is, at least partly, composed from things whose existence implies an awareness of them. As we saw in Chapter 1, there seems, at least, two ways of understanding this utterance. On the first way, the

things mentioned would be metaphysically dependent upon the state of awareness. But if the existence of these things requires the existence of a state of awareness for which the question once again arises how are we aware of this state, then no progress has been made. The postulation of these things provides just another way of launching upon an infinite regress. We would need to explain how the existence of such things was possible, and such an explanation would involve, in part, an account of how the subject is aware of them.

The second way of understanding the matter seems more satisfactory. We suppose that the things, whose existence implies that a subject is aware of them, are metaphysically sufficient for the awareness. They are such that if they occur in the mind of the subject, then the subject cannot help but be aware of them. In this case, the need for a state of awareness to explain why we are aware of them is not felt. It is true that we will be in such a state, but this will be a consequence of the occurrence of these things before a subject's mind. The important difference then is between a state of awareness that plays an explanatory role and one which does not. The state of awareness consequent upon the occurrence of the things whose existence is metaphysically sufficient for awareness is one which does not play an explanatory role.

It is obvious that the second way of understanding

the claim is the one we want. We want the things whose existence implies that a subject is aware of them to play the explanatory role and not the state of awareness. So, to summarise, the argument of the present section thus far is this.

- (1) If one is aware of something, then one must be partially aware of the nature of the state of awareness.
- (2) If the explanation of our awareness of something always requires that there exist another state of awareness, then there must be an infinite number of such states of awareness for subjects to be aware of anything.
- (3) There are not an infinite number of these 'explanatory' states of awareness.
- (4) Subjects are aware of things.

Therefore

- (5) It is not the case that the explanation of our awareness of something always requires that there exist another state of awareness.

Moving on to the positive part of the argument we get.

- (6) The only way in which the explanation of our awareness of something does not require that there exist an 'explanatory' state of awareness is if the existence of the thing in question is metaphysically sufficient for the subject, in

whose mind it occurs, to be aware of it.

Therefore, on the assumption that there is an explanation of our awareness of objects,

(7) There exist things whose existence is metaphysically sufficient for an awareness of them.

All we have to establish now is that the occurrences of phenomenal properties are the things mentioned in the conclusion, labelled (7). How can this be demonstrated?

One way to do this would be appeal to intuition. We may ask whether it makes sense to suppose that what we have identified as phenomenal properties could exist in the absence of a subject's awareness of them. Such an intuition, it might be thought, could have been consulted without all the hoopla contained in the present section. But by itself it would not have been enough to establish the desired conclusion. It is compatible with the occurrence of phenomenal properties being metaphysically dependent upon a subject's awareness of them. To establish their metaphysical sufficiency you need to consider how awareness is to be explained and the role phenomenal properties may play. What the intuition does do is soften one up to the use of notions such as metaphysical sufficiency and dependency and render plausible the claim that phenomenal properties do play such a role.

A second way in which it may be thought that one

could demonstrate that the occurrences of phenomenal properties are the things mentioned in the conclusion, rests upon the relationship between our account of phenomenal properties and any reasonable attempt to understand what it means to say that something has an existence which is metaphysically sufficient for a subject to be aware of it. For something to have that sort of character, one would expect that its full nature should be revealed in awareness. To allow otherwise would be to allow that there was some part of its nature of which the relevant subject was not aware, and, of this part, we would have to say that its existence was not metaphysically sufficient for a subject's awareness of it. But if the nature of something is entirely given in awareness then for that thing there will be certain distinctions between appearance and reality which do not apply to it. Appearance and reality will, in this special case, coincide. It might seem, therefore, reasonable to suppose that those properties which constitute the appearance of the mind to the subject, are precisely the properties mentioned in the conclusion. Phenomenal properties were specified to be such properties. Obviously there is leap here, but is the leap so wide?

Perhaps two things should be pointed out at this stage. The first is that although we said that for phenomenal properties, appearance and reality in some

sense coincide, it does not follow from this that we may credit ourselves with excessive epistemic privileges with regard to these properties. Their nature may compose the appearance of the mind, but who is to say we cannot be wrong about how the mind looks? The second thing we should point out is this. Even if the connection between phenomenal properties and our conclusion is not established, the conclusion itself indicates that there are some non-physical properties. So, at least the ramifications of this argument, for the argument for epiphenomenalism we are considering, are intact.

In so far as the leap mentioned a moment ago is legitimate, we have reason to believe.

- (8) The occurrences of phenomenal properties are those whose existence is metaphysically sufficient for an awareness of them.

This last proposition in turn entails the second of the two premises of the argument we offered at the beginning of the section. From which the conclusion follows that the occurrences of phenomenal properties are non-physical.

(c) What should we make of the argument?

In this section, we shall consider various objections to the argument just offered in the hope that in discussing them we will obtain a better appreciation of its strength or otherwise. Doubtless there are

objections we shall miss.

One objection to the argument may be this.⁶ We said right at the beginning that all awareness involved a partial awareness of the awareness. However consider once more the car driver. After many years of driving, it is possible, at times, to drive a car, as one might say, automatically. One is aware of what is going on in the road, but one's mind is also elsewhere. So, it is said, we are not aware of what is going on in one's mind in being aware of the road. Here is a case of awareness without awareness of awareness. The first claim of the argument of the preceding section is false.

In reply, we must note that it is important to distinguish between various orders of awareness, first order, second order and so on, and what one might call 'strengths of awareness' for want of a better phrase. Strengths of awareness may go from, at its weakest, just being aware as the car driver is aware of the road, to being aware of things as falling under certain concepts, to being aware in the sense of being inclined to make certain judgements concerning the object of awareness, to being aware in the sense of actually making the judgements, and so on. It seems quite possible that these various strengths of awareness may be reproduced for the various orders of awareness. The distinctions cut across each other.

Bearing this point in mind, the argument from awareness offered in the previous section concerned orders of awareness rather than strengths. It held that to be aware in even the minimal sense of the car driver, one must also be aware of the awareness and so on. The fact that the awareness is minimal is accommodated by the fact that the awareness of awareness is too. The same would hold for each awareness, as we further ascend the hierarchy. It is to be hoped that once this distinction is pointed out, the first claim can be seen to be true.

With this distinction in mind, it should be reasonably easy to predict how we are going to deal with a second objection, one stemming from David Armstrong's 'inner sense' theory of awareness. Armstrong's idea is that we are aware of states of our mind by having, effectively, mental sense organs that detect what internal states we are in, in a way analogous to the way our sense organs detect things in the world, and not because of the peculiar character of phenomenal properties. Thus, it may seem that our own approach is unnecessary.

Armstrong's position does not seem immediately susceptible to the infinite regress we described earlier because in response to the question, 'How are we aware of what we detect by the inner sense?' he will say that often we are not. When he says this, he means that often we are not aware of our awareness, and he thinks this claim is

plausible because he takes being aware of our awareness as involving a different strength of awareness. Understood in this way, his claim is quite correct. But if one does recognise the difference between strengths of awareness and orders of awareness, then it looks as if Armstrong's position is open to the objection canvassed in the preceding section. It results in an infinite regress of inner sense organs.

A third objection concerns the grounds of the claim that to be aware of something one must be partly aware of one's state of awareness, a claim which may be more perspicuously formulated as follows:

Metaphysically necessarily (if a subject is aware of something, then the subject is, at least in part, aware of the nature of his or her awareness).

Suppose, instead, that there is an awareness but the subject whose awareness it is, is not aware of the nature of the awareness at all. Then, the subject will be in no position to cite this awareness as a counterinstance to the thesis. The only instances that we can find of awareness will be, just because of the nature of our recognition of them, positive cases. So, perhaps the necessity that we feel attaches to the statement is just an illusion that has arisen as a consequence of the peculiarities of our epistemic position.

The obvious reply to this line of reasoning is to

recognise that whatever justification that we have for the modal claim stems from the conception that we have of awareness. It is by reflection on our concept of it that we recognise it could not apply to cases where there is no awareness of awareness. However, once we help ourselves to this reply we are open to a fourth objection.

Perhaps the argument we have offered has the same basis as the following. Suppose it is true that John steals a pencil from Peter. If John steals a pencil from Peter, then it follows that Peter is having his pencil stolen by John. We might even wish to say that there was a metaphysically necessary relationship between the first 'fact' and the second 'fact'. The reason why it is plausible that such a relationship does hold is that these putatively two facts are merely one. We have two ways of conceiving just one fact.

Now turn to the claim about awareness. May it not be the case that being aware of the nature of our awareness and being aware of the object amount to one and the same fact? In which case, it is no surprise that there is a metaphysically necessary relationship between them. We certainly need not suppose that there exist special non-physical properties to account for this situation. When we argued that unless there was an awareness of an awareness, and so on, there would not be an awareness, all this in fact amounted to was the claim that if there was

no awareness of the object then there would be no awareness of the object. Awarenesses are, so to speak, brute.

We may agree with the claim that an awareness of an awareness is no more nor less than a straightforward awareness. But, our reply should be that this does not solve the problem at which the line of reasoning behind the second premiss pointed. It was the question of what made us aware of anything. We suggested that calling something a state of awareness is just labelling the problem not providing a solution and that the only solution was to postulate entities that were metaphysically sufficient for a subject to be aware of them. Moreover, if this genuinely is the only solution, then no more detailed account of such entities will be able to get past the fact that they have a character which makes them non-physical. The psychologist, say, will just be providing more details about resolutely non-physical things.

A fifth objection concerns our attitude to the proposed existence of an infinite hierarchy of awarenesses. On the one hand, we claimed that there was not such a hierarchy of 'explanatory' awarenesses, on the other, we admitted that it was acceptable, in some sense, to talk of an infinite number of awarenesses in one. Is this not a contradiction?

The point that we were trying to make tacitly utilised the distinction, canvassed in the previous objection, between two conceptions of one fact and two facts. If a state of awareness is composed, in part, from phenomenal properties, then that state, by itself, constitutes the truth conditions of an infinite number of statements concerning the relevant subject's awareness, awareness of awareness, and so on. The occurrence of phenomenal properties, by being metaphysically sufficient for a subject's awareness, are sufficient for the truth of the statements that are a consequence of a statement of this awareness. These other statements are just conceptually distinct ways of recording this basic fact. What we said was unacceptable was an infinite number of facts of awareness. It was to this that we would be committed if we supposed that an infinite hierarchy of awarenesses was explanatorily necessary.

A sixth objection would be to challenge the claim that there should be a metaphysically necessary relationship between phenomenal properties and awareness. Why may the relationship not just be one of natural necessity? Of course, such a challenge presumes that one can make such a distinction, but let us allow that one can. The thought would be that given the way the laws are the occurrence of a phenomenal property would have as a consequence an awareness of that property by the subject

in whose mind it occurs. The problem with such an approach is that it locates the explanation for why we are aware of things in the causal relation. We are aware of the phenomenal property because it is causally related to a state of awareness. In which case, we must ask how are we aware of the state of awareness? And the answer will come back that it is causally related to another state of awareness, so resulting once more in an infinite regress. The point of claiming that the relation is one of metaphysical necessity is that it grounds the explanation for awareness in terms of the nature of what we are aware, namely phenomenal properties.

The seventh objection may be put this way. Instead of differentiating a putative state of awareness of a particular phenomenal property from the phenomenal property in question, we should identify them. In other words, we account for the existence of an occurrence of a phenomenal property implying an awareness of it by saying that the occurrence of the phenomenal property is the awareness. The difficulty that this suggestion has to face is that we have not concluded that phenomenal properties are non-physical because we cannot conceive of their existence implying that we are aware of them in any other way. It is not that we are looking for a way of making sense of this claim compatible with phenomenal properties being physical. Rather, we are faced with 'the

problem of awareness' and supposing that phenomenal properties are, in the relevant way, non-physical seemed the only way of dealing with this problem.

With this in mind, our question should be, if we account for the occurrence of phenomenal properties having their existence implying awareness by identifying the properties with awareness, do we retain the explanation of awareness that we previously felt we might have. The answer to this question is unfortunately no. The reason for this is that we now have no account of what makes a subject aware of the awareness. Our account had the virtue of explaining a subject's awareness by saying it was a metaphysically necessary consequence of the occurrence of the phenomenal property, the explanatory work being done by the phenomenal property. Once one identifies the phenomenal property with an awareness of it the explanatory work must be done by the awareness, and so the question about what makes us aware of the awareness becomes appropriate. The obvious answer is to propose that the awareness in question also has the feature the phenomenal property had, namely, to have as a metaphysically necessary consequence an awareness of it. However, then we have once more allowed the existence of something non-physical and hence just engaged in an exercise of renaming.

This brings us to a eighth and final objection that

must seem all too plausible by now. The problem, it may be said, with the approach we have adopted is that it has rested upon claims about a concept concerning which we have radical misconceptions, the concept of awareness.

It is extremely hard to criticise this proposal because the usual tools of criticism, if employed in the present case, are apt to be turned aside, their application being considered a consequence of the same misconception. So, the only effective criticism is a meta-claim. Here is one suggestion that seems to have something to recommend it.

The charge of misconception, if it is used to reject an argument, should be used sparingly. The problem with such a charge is that there are very few constraints, if any, upon the circumstances in which it can be made. As a consequence, it can underpin a rejection of almost any argument that results in a conclusion with which we are unhappy, or which clashes with some of our favoured beliefs. Therefore, we should only adopt such a charge if we can provide an adequate account of how somebody could be under such a misconception, and if we can show that the justification for their application of the concept that we say is misconceived is not only flawed, but also weaker than the justification for those beliefs with which the putative misconception clashes. Otherwise, we should accept the results of the application of the conception in

question. Both conditions are required since it is probable that there will be an adequate account of how we are misconceived for any of our concepts. So matters of justification must come in. On the other hand, until the account of our misconception is provided, the relative lack of justification in our application of a concept is not enough since the existence of a concept that has been systematically applied in the past requires some sort of explanation.

In the light of our constraints upon the charge of misconception, it seems fair to say that it is by no means clear that the charge as it concerns our concept of awareness is indeed respectful of these constraints. So, for the present, we should not recognise resistance to the argument from this area.

In this chapter we have tried to do two things. First, we attempted to understand what it would be to show that mental properties were physical properties. In so doing, we obtained an appreciation of the difficulty in establishing that mental properties were non-physical. Second, we presented an argument of our own for phenomenal properties being non-physical. If the approach in this chapter is correct we have discovered that either

phenomenal properties are causally epiphenomenal, or the physical world is not causally closed.

References

1. c.f. D. Davidson (1973), 'Material Mind' (in his (1980), Essays on Actions and Events), (1973), 'Radical Interpretation', Dialectica 27, (1974), 'Belief and the Basis of Meaning', Synthese 27; J. McDowell (1978), 'Physicalism and primitive denotation, Field on Tarski', Erkenntnis 13; D.C. Dennett (1979), Brainstorms, (1987) Intentional Stance.
2. Such a thought seems to be behind the insistence on topic-neutral descriptions in contrast to something more ontologically committing. J.J.C. Smart (1963), Philosophy and Scientific Realism (Ch.5), (1959), 'Sensations and Brain Processes' Philosophical Review 68; David Armstrong (1968), A Materialist Theory of Mind (Ch.6) For a qualification to this position, see D.M. Armstrong (1973), 'Epistemological Foundations for a Materialist Theory of the Mind', Philosophy of Science.
3. e.g. S.A. Kripke (1972,1980), Naming and Necessity; T. Nagel (1974); F. Jackson (1982), 'Epiphenomenal Qualia', Philosophical Quarterly 32, (1986), 'What Mary didn't Know', Journal of Philosophy 83.
4. G. Ryle (1949), The Concept of Mind (Ch.7). Ryle's remarks applied to the question of whether sensations were ascribed given that all observation involves sensation, but it is clear his remarks have application here too.
5. The issue was raised by Aristotle, see De Anima (Bk3 s2). We shall assess Aristotle's answer in the next paragraph.
6. See footnote 2 for references, and, D.M.Armstrong (1980), 'The Nature of Mind' (in his, The Nature of Mind).

Chapter 7

On the nature of intentional properties

The aim of the present chapter is to assess whether intentional properties have physical occurrences, or the predicates ascribing intentional properties, physical conditions of application. One way in which neither of these things is likely to be true is if intentional properties turned out to be phenomenal properties, or to be wholly or partly composed from phenomenal properties, because, then, the conclusion of the preceding chapter may have application to them. Of course, if they were only partly composed from phenomenal properties, the other components could be causally efficacious, and so, intentional properties could stand in causal relationships.

Since our concern is not specifically with the character of intentional properties, but with whether or not they are epiphenomenal, we shall not try to assess whether they are related to phenomenal properties in the way suggested.¹ Our further discussion of the explanatory role of phenomenal properties would just have more extensive application if intentional properties were so

related. As our overall aim is to find out something about the role of mental events in psychological explanation, and not to come to interesting but inessential conclusions about intentional properties, we shall spend our time considering various accounts of how intentional properties may have physical occurrences or the predicates which ascribe them have physical conditions of application. As the accounts which we consider do not, in general, appeal to this distinction we shall hereafter drop it, and just talk non-committally of whether an account can be provided of intentional properties in physical terms.

Our overall conclusion will be that it is possible that such an account can be given, but we must note up front two issues which cut across it. The first matter concerns our analysis of the physical. It may be felt that the various inadequacies of the accounts of intentional properties that we consider reveals an inadequacy in our analysis. According to it, to be non-physical, the occurrence of an intentional property would have to imply that a subject is aware of it. But, not only is this unlikely to be true, it is also unlikely to help us understand the character of intentional properties. So, intentional properties are physical. Yet, someone might say, intentional properties are obviously non-physical and our failure to recognise this

results in paying needless attention to a lot of hopeless theories.

In reply, we should note that the account of the physical we adopted was deliberately non-committal over whether phenomenal and intentional properties were physical. It took an argument to establish that phenomenal properties were. We should not just assume that intentional properties are non-physical. Moreover, what purchase do we get upon our problems if we take up such an option? Calling something non-physical does not solve anything and it was not our intention to make 'non-physical' synonymous with 'mysterious'.

A second thing upon which we should remark is our use of the phrase 'means of representation'. It has been chosen precisely so that we may remain untroubled over whether humans have intentional contents because they have a 'language of thought' or whether such a language is unnecessary.² It is to be hoped that this ambivalence allows us to focus entirely on the various accounts of intentional properties given below. Of course, it may be that one or more of them could only hold of minds if their means of representation had a particular character, but the type of objections we bring forward do not depend upon inadequacies in the means of representation, so we may ignore this issue.

(a) On the problem with assessing any account of intentional properties in physical terms.

The basic difficulty with assessing any account of intentionality is that many philosophers hold that certain attributions of intentional content are illegitimate on the basis of their being unaccommodated by the account they offer.³ In each case, their account of what intentional contents are, or of the facts upon which they depend, suggests to them that certain intentional contents which we would otherwise take to exist, in fact, do not. The problem for those who wish to assess their accounts is that there is consequently no generally agreed evidence.⁴ The inability of an account to accommodate a certain type of intentional content, instead of being considered as a sign that the account is inadequate, is taken as a sign of the content itself being illegitimately ascribed. So, one is at a loss to tell what would count as an objection to the accounts in question.

The situation just described has another consequence. It would be inappropriate to apply the procedure we followed with regard to phenomenal properties, to intentional properties. There is no point in first arguing for the existence of intentional properties and then considering accounts of their nature. The accounts offered are supposed to motivate conclusions as to the existence or otherwise of intentional properties.

Consequently, we shall just consider the accounts and skip the initial stage.

It is important to recognise one reason for the present situation. When one provides an interpretation of a means of representation one attributes to elements of the means of representation, intentional properties. Interpretation of languages or even, perhaps, works of art have individuals who are to some extent authoritative on what intentional properties they possess. For instance, there are a collection of competent language users who know the language and can teach it to others. However, this situation is not reproduced when we consider providing a physical account of intentionality. The properties which compose our brain are not invested with their significance by intelligent users. They are not anybody's language or means of representation. They have their intentional properties because of their character alone, and thus nobody has a special claim to be able to read that character.

The problem that we are faced with can seem intractable and therefore encourages a certain sort of intellectual behaviour. Suppose one has a moderately successful account of intentionality on one's hands. There will be certain sorts of intentional properties that we can detect by using it, but others that we took to exist, which we cannot detect. Since it is difficult to

ascribe or understand intentional properties in the first place, and since the account has been so useful, one is apt to think that the account cannot be wrong, but, rather, it is the prior belief we had in the existence of certain sorts of intentional properties which is. The evidence for their existence may be dismissed as uncertain. Thus, certain ascriptions of intentional contents to subjects that we intuitively would make, are rejected as illegitimate.

We shall try to avoid the debate over which intentional properties exist by making only one principal assumption: that it is possible to make mistakes of judgement. Various accounts will be considered solely with regard to their ability to deal with this fact. Our discussion of each will be brief. The aim is to develop a general line of criticism that will result in a diagnosis of the root of the problem. This will then serve to motivate our positive proposal. Thus, we shall not consider in any detail possible improvements of the accounts of which we are critical, which would otherwise be appropriate. Instead, we may claim that any such improvements should lie within the overall format we propose at the end.

(b) On the reason why intentional properties cannot be taken as primitive

Each of the accounts of intentional properties that we shall consider below attempt to help us understand the character of these properties. But, it may be thought that consideration of them is unnecessary. Our preliminary description of intentional properties as 'capacities or sub-capacities to refer' indicates that these properties are not non-physical. Nothing in this description suggests that the existence of things satisfying it imply a subject's awareness of them.

However, to stop short at this point would be unsatisfactory. We have, as yet, little idea of what it is to possess such capacities. Thus, it may turn out, upon further scrutiny, that the capacities are non-physical. Also, in considering their character, we will be laying the groundwork for the second challenge of epiphenomenalism.

(c) The dispositional theory of intentional properties, and the problem with making mistakes.

Suppose we wish to provide an account of what it is for a subject to possess a certain concept, one type of intentional property. One way to begin is to chronicle the judgements that the subject would be able to make if he or she had that concept. Suppose the concept is that

of a dog. The judgements the subject would be able to make would be those concerning dogs, assuming that he or she possesses the other concepts relevant to the judgements. We only make progress, however, if we can say what it is about a judgment that makes it of a dog, rather than something else. Since we have tried to account for what it is to possess a concept by citing the judgments a subject would be able to make if he or she did possess the concept, we cannot say that the judgements are judgements of dogs because they involve in some way the concept of a dog. The alternative suggestion that has seemed plausible to many is that a judgement is about a particular thing if it is responsive to evidence of how things are with that thing in the way relevant to the judgement. A judgement is about a dog if we are inclined to make that judgement as a result of evidence concerning dogs, or, if we have made it, inclined to view the truth or falsity of the judgement as something which may be assessed on the basis of such evidence.

We may call the rudimentary account we have just put forward a 'dispositional' or 'functional' theory of intentional content. According to it, types of means of representation have intentional properties when they occur in a subject's mind, if each type occurs in the 'appropriate' circumstances. Thus, for instance, an occurrence of a means of representation has the

intentional property of being the concept of a dog if it is the occurrence of a type of means of representation that would occur when there was evidence available concerning dogs.

We said, at the beginning, that any account of intentional properties should be able to deal with the fact that errors of judgement can be made. So the question is 'Does the theory before us legitimate the ascription of certain intentional properties that will allow for this fact?'

Suppose I am faced with a cat dressed up as a dog. I judge that what is before me is a dog, an error. But why should we suppose that I be subject to self reproach in this way? Why not ascribe to me the concept expressed by the predicate "- is a 'dog'" where:

X is a 'dog' if and only if

(a) It is not the case that X = a cat dressed up as a dog, and X is a dog

or (b) X = a cat dressed up as a dog, and X is a cat dressed up as a dog?

In such a way, we can eliminate all errors.⁵ Since the ascription of these types of intentional properties seem to be more responsive to the circumstances in which some type of means of representation occurs, it would appear that their ascription would be more appropriate than the intentional properties that allow for error.

One might try to get past this difficulty by suggesting that our disposition to correct our judgements and recognise errors as errors enables us to legitimately interpret my concept as of a dog and not a 'dog'. However, it is not obvious how the problem is resolved by adopting this approach. It is an unfortunate truism that we tend to make errors as to what our errors are. Therefore, not only will we be unable to filter out all our mistakes but also a similar question arises as to whether I indeed possess the concept of error. Why should we ascribe to a subject the concept of an error rather than the following concept.

X is an 'error' if and only if

(a) It is not the case that X = one of the situations in which we miscorrect ourselves, and X is an error.

or (b) X = one of those situations in which we miscorrect ourselves, and X is not an error.

It is arguable that other attempted solutions to this problem only succeed by assuming that the preferred ascription of a concept over another is legitimate.⁶

We can fail to make correct judgments in other ways. Presumably, we would be correct to judge of the sun's surface that it is hot, yet, we would not make this judgement if we were actually upon the surface. The question is, why should we ascribe to ourselves the

intuitive concept of being hot rather than one which has the same application as it except that it does not apply to the sun when we are on its surface? Of course, it is tempting to think that there is some mental structure which, if it were not ensconced inside flesh but in more heat-resistant material, would enable a subject to make the relevant judgement on the sun's surface. The problem with such a proposal is that it is unclear why in the case of concepts people should be allowed to have capacities they cannot in fact manifest, but in the case of strength, say, they are not said to possess the capacity to lift a car because a suitably reinforced version of a human body would be able to.

To deal with this problem, what we seem to need is some way to differentiate between those occurrences, or failures to occur, of some type of means of representation which should not be counted towards determining what intentional property it possesses, and those which should be counted. Roughly speaking, there are two possible responses. One might try to cite some 'internal' fact about a subject's mind which alone differentiates between these two circumstances. Alternatively, one might try to appeal to some external fact that makes the difference. The accounts that follow fall in one of these two categories. Some are meant to supplement the

dispositional account. Many by providing the requisite difference aim to supplant it.

(d) **Will the notion of resemblance have any role to play in a physical account of intentional properties?**

One proposal that is sometimes put forward is that a type of means of representation represents an item in the environment if it resembles it. The most plausible version of such an account will rely upon an independent story about what makes a means of representation a means of representation to take care of a standard objection against such theories, namely, that if resemblance is going to do all the work, the item in the environment represents the means of representation as much as the latter represents the former, since resemblance is a symmetric relation. If we further remember that we need only count resemblance as a sufficient condition for representation, other objections are avoided.⁷ Nevertheless, the theory still faces what look to be serious problems.

One problem stems from the fact that it is implausible to require that the resemblance between means of representation and the item which is represented must be total. But, if resemblance in certain respects only is required, we should want to know what it is about the workings of the representational system that determines

that it is the resemblance that matters rather than the differences. The danger is that we can only make sense of a resemblance being taken as important if there is something like a perceiver of the resemblance. For example, a cartoon of a famous figure is very unlike the famous person, but similarities are seen by those who inspect the cartoon. However, if we take this way out it seems that we have not accounted for intentionality since to see something as similar to something else, and therefore representative of it, must be to perceive that the representation has such and such features, and judge that these features are similar to the features a thing has. But, each of these involves intentional contents.

It might be thought that an appeal to the dispositions of a subject's mind could resolve the problem concerning which aspect of a means of representation should resemble what is represented and that an appeal to the resemblance relation could resolve the 'problem of error' described in the preceding section. The problem with this idea is that it is unclear whether these two aspects are compatible.

Once one admits that the dispositions a subject has with regard to a particular means of representation determines how the resemblance relation becomes a relation of representation it seems that the argument of the preceding section can be reproduced. Any occasion in

which we may be tempted to say that a mistake has been made could instead be taken to be evidence that the resemblance that we thought determined what was represented in fact did not. Why should we suppose that what matters to the mind, so to speak, is the resemblance if it is disposed to overlook aspects of the resemblance? Nor can this problem be avoided if we suggest, rather implausibly, that some types of means of representation resemble what they represent in every respect. The question still arises, why should this resemblance be counted as determining what is represented when the mind, so to speak, ignores its pronouncements?

The objection we have offered highlights the problem we sketched in the first section of this chapter. The resemblance of the means of representation to the item to which it makes reference renders it easy to read its intentional properties. It is as if what represents has been constructed by a human author, and accordingly imbued with meaning. However, this resemblance may be of no significance as far as the intentional properties of the means of representation are concerned. To determine whether it is, we cannot just look with our own eyes and see what significance the means of representation has for us. We have to consider its significance, so to speak, for the mind whose means of representation it is, and once

we do this, it may seem as if the account of intentionality has to come from elsewhere.

(e) **Will a causal relationship between what is referred to and the means of representation account for intentionality?**

It has been thought that the existence of some type of causal relationship between a means of representation and an item in the environment suffices for the former to represent the latter. Once more we will need an independent account of what a means of representation is, otherwise every causal relationship of the relevant kind will involve representation. We should hold that representation is only present when one of the relata is a means of representation.

It is pretty clear from our discussion of the dispositional theory of concepts that the following proposal is inadequate, (using 'item' to be neutral over what kind of thing is represented).

A means of representation of a certain type represents the item in the environment with which it is causally correlated.

In fact, the proposal is just a version of the dispositional theory with the emphasis placed solely upon external relations. It is therefore susceptible to some predictable problems, most of which have been well

documented elsewhere.⁸ It is also obviously susceptible to a version of what we have called the problem of error, and which Jerry Fodor has called the disjunction problem.⁹ According to the account specified above, so long as a means of representation of a certain type is causally correlated not just with what it is supposed to represent, dogs say, but also with something that it is not, cats dressed up as dogs, then that means of representation, in fact, represents the disjunction we specified, either dogs, or cats dressed up as a dogs. So, once more we are looking for some fact which differentiates those occurrences of a means of representation which do determine its intentional properties from those which do not.

One attempt to separate the causal correlations which confer upon a means of representation its intentional property from those which do not can be extracted from Fred Dretske's work.¹⁰ Simplifying a little, but not in a way which will affect the point that we will make hereafter, we may take his proposal to be that:

A means of representation of a certain type represents an item in the environment if during the learning period for it, the following association is set up: the means of representation of the type indicated is causally correlated with the item.¹¹

The way in which this account solves the problems we

described above is by allowing that a certain means of representation may be caused by other items than those which it represents after the learning period. The intentional property that the means of representation has is determined by the association set up during the learning period and is more or less unaffected by what happens thereafter.

There are some difficulties with the account.¹² One of them is that it assumes that eventually the learning period comes to an end. Yet, it seems plausible to think that learning how one should correlate one's means of representation with items in the environment never comes to an end. For example, one might think that we have come to the end of the learning period of that means of representation which stands for dogs. Suppose a new type of animal is now found in some remote region of the world which, according to the criteria we established during our putative learning period, both counts as a dog and does not count as a dog. Two apparently sufficient conditions for doghood and non-doghood have come apart. Arguably, in these circumstances, we would have to conclude that we had not formed an adequate concept of dogs and, thus, the learning period of the means of representation was not over. Therefore, the false thoughts we had hitherto ascribed to ourselves about dogs could not really be false for the means of representation

we used to think them lacked intentional properties.

Another flaw in the account concerns how, in the learning period, a particular association is set up. Dretske assumes that during this time the learning subject is in optimal circumstances. It is difficult to see why this should be so. The learning period is bound to contain occasions when the learner 'makes a mistake', for instance, when he or she is required to perform a difficult discrimination essential for learning about a particular thing. Since the means of representation of a certain type represents with whatever it is correlated in the learning period, it cannot plausibly represent what we take it to represent. The concepts we have of things are not concepts of things about which we made no mistake in the learning period.

One might think that the existence of a teacher during the learning period would help make, what would otherwise be, sub-optimal circumstances optimal. The teacher would correct mistaken correlations in the learning period and make the object of learning salient to the pupil, and, thereby, attempt to set up the right associations. However, there are problems with this suggestion. First, it is hard to see why the means of representation, which the subject would then produce, would not represent the correction by the teacher as well as the item that it is supposed to represent in the

circumstances in which a correction is necessary. Second, the teacher will make mistakes in his or her corrections and in the objects which he or she makes salient. Yet this, arguably, does not always affect what is learnt.

Another problem with the account may be put this way. If the association set up during the learning period covers the application of the concept in all relevant circumstances, then it is hard to see why a subject should thereafter make mistakes. Is this a result of 'unlearning' and, if it is, should we not take it to have consequences upon the intentional properties correctly ascribed? If, on the other hand, the association set up does not cover the application of a concept in all relevant circumstances, then an alternative specification of what is learnt may be given in terms of a concept in which all the relevant circumstances have been covered. In which case, the question is, what fact about the learning period rules out the latter concept being the one learnt? This is just a more specific version of the question that, in general, this account was meant to answer.

The second attempt that we shall consider to provide a theory of representation based upon causal relations is due to Jerry Fodor.¹³ It not only is the most advanced account yet offered of this sort, it also reveals what may be thought the basic problem with the whole approach.

Fodor's account is this:

A means of representation of a certain type represents an item in the environment if

(i) The means of representation is causally correlated with the item in the environment in optimal circumstances.

(ii) Any other causal correlation which holds between the means of representation and a different item in the environment only holds because of the correlation mentioned in (i) and not vice versa.

In putting forward this account, Fodor has basically offered something akin to one of our ways of dealing with the problem of spurious causation. However, since he allows that a causal relation does exist between the unrepresented items and the means of representation, it cannot be that the spuriousness stems from the lack of a causal link. From where does it stem?

Let us return to our example of dogs and cats. Fodor claims that although there may be a causal correlation between the means of representation of dogs and either dogs, or cats in dogs' clothes, the reason why the means of representation of dogs is of dogs is that the causal correlation between cats dressed up as dogs and the means of representation only holds because dogs and that means of representation are causally correlated, and not vice versa. But how is the causal correlation between cats

dressed up as dogs and the means of representation of a dog dependent upon the other causal correlation? It is in trying to answer this question that Fodor's account becomes problematic.

The natural answer to give is that cats dressed up as dogs look like dogs. So, we make mistakes. Our propensity to correct ourselves, when we do, makes the causal correlation between the represented item, dogs, and the means of representation the independent causal relation, and the other causal correlations dependent upon it, by being less 'robust'. But, if it is just the subject's disposition to correct him or herself that establishes the independent causal correlation, and renders the others dependent, then the arguments contained in section (c) of this chapter have application to the present account.

Fodor¹⁴ deals with the problem just mentioned by suggesting that there is some other fact about the subject that results in the 'asymmetric dependence' of one of the causal correlations on the other. According to Fodor, this fact make certain possible worlds more similar to the actual one than others and, thereby, constitutes the required dependence. If we consider different possible worlds in which the subject is psychologically similar, and the same laws of nature hold, then the possible world in which the relevant type of means of representation is

applied, by the subject, only to dogs, is more like the actual world than the possible world in which it is applied only to cats in dogs clothes.

Suppose we grant that this is so. Unfortunately, this still does not seem to give us an acceptable account of error. A possible world in which 'errors' are made, and thus, the type of means of representation is correlated with what we could categorise as a disjunctive property, will be even more similar to the actual world than the one in which the correlation is only with dogs. The causal correlation between dogs and the means of representation is, consequently, dependent upon that between the putative disjunctive property or, perhaps, what is common to the disjuncts, and the means of representation. Thus, we should take the concept to be of one of these.

Of course, Fodor could rule out such worlds, but not only is this unmotivated, it may in fact rule out intuitive assignments of intentional properties. We should not presume that what we conceive of as disjunctive or not disjunctive is in fact this way. The property of being a dog may be disjunctive. So, even if we allow that Fodor's account can make sense of there being an error of judgement, the achievement appears minimal. If we overlook the actual world, then the most similar possible world is probably one in which the 'disjunction' with

which the relevant type of means of representation is correlated differs only slightly from the pattern of judgements in the actual world. At the extreme, Fodor might be left with the assignment of an intentional property that allows one case of error.

(f) Can we provide an account of intentional properties in terms of the function of the things which possess them?

The principal idea behind the range of approaches we are now going to consider is this:

A means of representation of a certain type represents an item in the environment if the means of representation has as its function to so represent the item.

If something has a certain function then, crudely, we shall suppose that it:

- (a) causes a certain effect, or range of effects,
- (b) exists in order to bring about that effect or range of effects.

It is the second condition which requires most elucidation. What we must do is, first, identify a 'function-conferring' set of facts, and, second, establish that the function-conferring facts attribute functions that allow for errors of judgement.

In a revision to the account he offered earlier,

Dretske suggests that¹⁵

A means of representation of a certain type has as its function to represent an item in the environment if

(i) The means of representation of that type was causally correlated with the item in the environment.

(ii) The means of representation of that type has an affect upon the behaviour of the organism for which it is a means of representation because of (i).

Although it is not explicit in our formulation of Dretske's current account, once more the key to intentionality is seen as learning, in particular, discriminative learning. Suppose the occurrence of some internal component of an organism is correlated with the occurrence of a certain item in the environment. Suppose further that the occurrence of the internal component on occasion gives rise to a piece of behaviour which is successful because the item in the environment is present. The success of the behaviour has certain causal consequences. The organism gets 'rewarded' and this, in turn, has the consequence that the internal component is more likely to give rise to the piece of behaviour whose success relies upon the discrimination of the relevant item. It is in this way that Dretske envisages (ii) is

satisfied. The means of representation possess a function, and, thereby, an intentional property, because it exists in order to have a particular effect upon an organism's behaviour.

Dretske would be the first to admit that the account just offered is sketchy. So, any discussion of the details of its application is bound to be inconclusive. The thought is that if we grant that the means of representation has the function of indicating the presence of an item in the environment because it has been 'selected' to play a role in a subject's behaviour, then it may on occasions occur when it should not, and thus misrepresents the world. It would retain its function in these circumstances because it still plays the role in causing behaviour for which it was selected. However, for it to have the function to represent an item, it must have been perfectly correlated with the item in the environment throughout the period of time in which it was enlisted, and this requirement raises the same problems we found with regard to the learning period of Dretske's previous account. The problem just seems to have been relocated. Until one can resolve this matter, it would not be legitimate to suppose that some occurrences of means of representation that give rise to behaviour occurred when they should not.

Recently, it has been argued, by Ruth Millikan and

David Papineau, that the notion of natural selection, and evolutionary theory in general, can provide insight into how the means of representation obtain their intentional properties.¹⁶ Natural selection is taken to supply us with a scientifically acceptable notion of function, and means of representation by possessing functions of this sort have intentional properties.

We are used to the idea that something may possess a function that is invested in it by a designer. If God had constructed the human brain so that certain components were supposed to occur when dogs were present to the senses, then those components would have had the function to represent the occurrence of dogs even if they did this only imperfectly. Evolutionary theory is supposed to provide another way in which we may understand the possession of a function by certain things, one which does not appeal to a designer. It is suggested that nature is so constructed that it acts as a selection mechanism allowing certain things to proliferate and others to occur less frequently. The selection occurs with regard to living organisms and their features. Most living organisms have the potential for reproduction. The definitive character of natural selection is that it occurs as a result of certain things being able to reproduce or not reproduce.

Reproduction involves both replication and variation.

The latter is necessary for evolution to take place, the former is crucial to understanding function from the evolutionary perspective. Suppose there are certain features an organism may have which will make it more likely to be able to reproduce. It would not follow that this was their function. Equally, suppose that reproduction resulted in the production of something completely different, for instance, a badger might give birth to a three-piece suite. Again, the notion of function would have no role to play. What is of interest is the case where something similar is produced, and, in particular, some feature of an organism is produced in the offspring of that organism because that feature was partly causally responsible for the ability to reproduce.

Crudely put, the idea is that if a certain feature of a creature is reproduced in its offspring because the creature produces creatures more or less like itself and the feature has helped the creature to be able to reproduce, then that feature has a function. More precisely, Millikan argues:

A feature has as a function the playing of a certain causal role if

- (i) Previous occurrences of the feature tended to play the causal role.
- (ii) The tendency to play the causal role caused the reproduction of that feature.

There are additional complications in Millikan's account but they need not concern us here. We shall consider two questions. First, is the account just presented an adequate account of function? Second, can it be utilised in an account of intentional properties?

The following seems to be a counterexample to Millikan's approach. Suppose that I am nervous whenever I think that I am in the company of people brighter than myself, and, that my nervousness makes me more likely to say foolish things which makes the people, in whose company I am, more prone to be condescending. Being condescended to on these occasions makes me more likely to be nervous in the future, in such circumstances. Still, one would not want to conclude from this that the function of my nervousness was to make people condescending to me. Yet, it would seem that one must conclude that this is so according to Millikan's account.

The unintuitiveness of counting my nervousness as having the function indicated would seem to place a question mark over Millikan's account of function as it stands. However, perhaps we have failed to bring out a crucial feature of her account, namely that we should consider an entity to have a function if it has a role in the achievement of some 'scientifically significant' end, such as reproduction. In which case, the following addition should be made to the account:

(iii)The effect that the occurrence of the feature tends to bring about, in playing the causal role, is judged to be scientifically significant.

We shall not consider what it is to be a scientifically significant end, although we shall assume that reproduction is such an end. An interesting feature of this revised approach is that it leaves open the possibility that other functions may be defined in terms of other scientifically significant ends. Had it been the case that a successful account of function was provided with only conditions (i) and (ii), then Millikan would have been in a position to point to a unique collection of function-conferring facts, those of natural selection, which could legitimate her claim that intentional properties should be understood in terms of it. Instead, we see that a function specified in terms of reproduction is one out of a range of scientific perspectives we might adopt.

With this qualification in mind, let us consider whether Millikan's notion of function can be used to provide an account of intentional properties. We may abstract our discussion from the different stories that Millikan and Papineau provide as to how the selection of means of representation takes place, and from the matter of whether what is selected are individual means of

representation or systems producing such means. Our criticism will not rely upon such issues.

One objection is that, according to this approach, as both Millikan and Papineau admit, a means of representation can only have an intentional property if the organism possessing it has a reproductive history. As they recognise, it follows from this that if, by some cosmic accident, a man was formed from a dust cloud, which was exactly like me in every physical, phenomenal and dispositional way, that man would still have no means of representation with intentional content. Yet, one has the intuition that the man in question would not lack such content.

Both Millikan and Papineau reject the objection canvassed above by denying that what is represented is something to which we have immediate access. Both rail against what they characterise as 'meaning rationalism', the belief that intentional content is something given to us which we will believe to be possessed by a means of representation just when it is. Naturally, it would be unwise merely to claim that such and such a thing is 'given to us', this could seem like prejudice. Nevertheless, we can fairly say that the similarity between my twin and I gives us some reason to believe that we both have intentional properties. Equally, the fact that if I were the dust-cloud man, I would take myself to

have intentional properties is some evidence that I would have such properties.

In fact, there is a potential clash between meaning rationalism and the evolutionary perspective. One would expect that if intentional properties record the biological utility of a particular means of representation then an ability to detect the presence of intentional properties would itself be favoured by natural selection. So, if it is correct to adopt the evolutionary perspective in this area, one should, in general, adopt a weak version of meaning rationalism. In which case, rejection of the latter cannot be used to defend the evolutionary perspective.

Even if we overlook the objection just made there is another which severely threatens the plausibility of the account of representation on offer.¹⁷ The account suggests that the intentional property which revealed why a means of representation was conducive to the reproductive ability of an organism should be assigned to that means of representation given the right reproductive history. But there seem various ways in which one may do that. Consider the disjunction problem we discussed earlier. We shall assume that in certain circumstances it is possible to mistake an old English sheep dog for a tiger. We shall further assume that tigers are dangerous. Suppose that when faced with a tiger or an old English

sheep dog in bad light, I engage in what one might colloquially call 'evasive action'. If I had been faced with a tiger the action would have been appropriate, otherwise not.

On the assumption that my behaviour is typical of the behaviour of those who make up the selective history of organisms of my type, one might argue that the intentional property ascribed to the means of representation that causes my behaviour should be the concept of tigers rather than tigers or old English sheep dogs in poor lighting. It is the ascription of this intentional property which would show how my behaviour was conducive to my survival and therefore my reproductive ability. However, one could equally ascribe the concept of tigers or old English sheep dogs in poor light and my behaviour, and the behaviour of those in the past, could still be shown to be conducive to my, and their, survival. The means of representation with that intentional property would result in behaviour which is more often than not conducive to survival in the world in which we live so long as the means of representation is correlated with tigers. We do not have to invest all the utility to survival in the content itself. Thus, it looks as if Millikan can provide us with no account of misrepresentation. It appears that the point can be generalised. All errors can be eliminated by speaking,

instead, of the appropriateness or inappropriateness of behaviour. So, unless something can be provided to motivate the location of the utility to survival in the means of representation, itself, it seems that Millikan and Papineau have no non-arbitrary resolution of the problem of error.

(g) **A tentative positive proposal.**

It has been argued by a number of philosophers recently that¹⁸

Some mental events are metaphysically dependent upon the environment for their existence.

We shall not go into their arguments here. Suffice it to say that they have claimed that we have a body of intuitions regarding our ascription of intentional contents of certain kinds which makes the correctness of the ascription dependent upon the existence of features of the environment, although, not necessarily the environment at the time of the ascription. The environment identified includes the natural world, as well as the social and linguistic environment. The intentional contents ascribed are taken to be genuine constituents of some mental events and it is these mental events that are responsible for the truth of the thesis above. The claim of metaphysical dependence is supposed to be one concerning the nature of mental events with these intentional contents.¹⁹ For

example, it has been suggested that some subjects could not have the belief that water is wet unless they had come across water in the environment. Only those who could identify water independently of the environment could have such beliefs even if they had not come across water, because they could single out water for themselves, in thought.

What we shall do is suggest that the perspective on intentional contents introduced by the recognition of this thesis should make us more optimistic about the possibility of resolving the central problem discussed in this chapter, despite the negative conclusions to which we have so far come. It should not be thought that of all the accounts of intentional properties we have considered, none have the upshot that intentional contents are environment-dependent. One may reasonably claim that most of the accounts examined had the consequence that, at least, some intentional contents were. However, by our lights, they did not exhaust the range of possibilities open to one, once one recognises the fact that some intentional content is so dependent.

To begin with, we should recognise a distinction between two ways in which one might say that a subject has a conception of some entity, or some type of entity. According to the first way, for a subject to have such a conception, the subject must either have

phenomenologically before the mind the wherewithal to single out the entity or type of entity of which the subject has the conception, or there should be, over some significant period of time, a causal correlation between occurrences of a type of means of representation and the putative object of the subject's conception. To put the point in a line, what determines the object of conception is 'subject-centred'.

The alternative approach that we shall adopt places the emphasis upon the items in the subject's environment. The idea is that we partly rely on the items to make a conception, a conception of them. Of course, it would not do to just single out any component of a subject's mind, and say that it is a conception of an item in the environment. One needs to provide some reason for thinking that the component should be counted as a conception of that very thing. But, the reason supplied need not be of the demanding form that 'subject-centred' accounts suppose it to be.

Instead, the account takes the following form. It suggests that there are certain facts that constitute a subject making cognitive contact with an item in the environment, for instance, by being perceptually aware of that item. Suppose that, at the time in which cognitive contact was made, a certain type of means of representation was associated with the item, for instance,

by the means of representation's occurrence being a causal consequence of the contact. It is then claimed that there are other facts about a subject's mind which, should they hold, mean that the means of representation when it now occurs, refers to the item, or to an item of the same type, with which the means of representation of that type was originally associated, at the time of the cognitive contact. One may take the accounts offered above in terms of dispositions to make judgements and causal relations as attempts to fill out our understanding of these facts, but these ways of filling out our understanding are not compulsory.

The alternative is to recognise that we need far more empirical research before we are likely to obtain a precise idea of the nature of the facts in question. It is important to be clear about how this proposal is going to work. The thought is that we have already a clear idea of the circumstances in which it is correct to describe a subject as having made cognitive contact with some item in the environment. Equally, we have some idea of when it is appropriate to ascribe intentional contents to subjects. If we could also come to an independent understanding of what may be the means of representation of a mind, then we could use the materials at hand to find out how the means of representation would be associated with a certain item via the making of cognitive contact, and what facts would

still have to hold, and what not, in a subject's mind, for the means of representation to retain its representation of that item. The suggestion is that we take seriously our actual attributions of intentional contents and use our most stable ascriptions and the evidence upon which these are based, to determine in detail the character of the facts we sketched in broad outline above.

Resistance to this proposal may come from a number of quarters, but part of this may be based upon what, in our view, is a misapprehension of what the character of the facts mentioned must be. It is often thought that the type of facts cited should be such that one sees by inspecting them that they are such as to constitute part of the occurrence of intentional properties. When one is first presented with a version of the causal account, prior to the presentation of the counterexamples, one feels a strong intuition to say of a certain means of representation that it must have the required intentional property because of the causal relation which holds between it and some item in the environment. But, upon our view, this is to make the same mistake as that which made the resemblance account seem so plausible. Instead, we should not suppose the facts to be intuitive components of some occurrence of intentional properties. Everything about them is still to be discovered.

To try to illustrate further the approach suggested,

we shall consider briefly four features of it. The first is that it depends upon the existence of a reality which is undetermined by our conception of it. Instead, reality determines the character of our conceptions since we utilise constituents of reality to demarcate the character in question. For a means of representation to have an intentional property which is that conception is just for it to stand in a certain relation to that constituent of reality, a relation which is not shared by other means of representation with different intentional properties. In noting that this is a feature of our account, we should not be taken to complacently suppose that this notion of reality is uncontroversial. It is not. However, the theories of intentional properties we discussed usually presumed that such a notion was legitimate. So, we can view our approach as presenting an alternative for those motivated to put forward the accounts we considered. In a more comprehensive discussion of intentional properties, it would naturally be of interest to examine the relationship between the conception of reality one adopts, and the type of problems one faces in providing a theory of their nature.

The second feature of the approach that we should appreciate is the way it deals with the problem of error. The advantage it brings, by moving away from the subject-centred approach, is that it separates the judgements a

subject is inclined to make, or the pattern of occurrences of types of means of representation through time, from the intentional properties that these means of representation possess. It was this emphasis that gave rise to the problem of error since the pattern mentioned, as a result of error, never seemed sufficient to determine an intentional property that allowed for error. By locating the facts that determine the intentional properties elsewhere, the problem does not arise and errors can once more be taken at face value. There is, after all, no reason why there should be a connection, between the pattern of occurrences of a means of representation, and its intentional property, once one rids oneself of the idea that the facts in question must be intuitive components of an account of intentional properties, rather than just those things we discover as a result of empirical research.

A third feature of the approach, that we should recognise, is that it is compatible with there being no single condition that a subject must meet for him or her to be correctly ascribed an intentional property. For instance, suppose that the subject satisfied the dispositional theory with regard to the possession of the concept of dogs by, in the long term, never making an inappropriate judgement concerning dogs. Any error was corrected. In such a case, we may allow that the

satisfaction of the dispositional approach is sufficient for an attribution of the intentional content, but only because there is the other approach to intentional contents that reveals how a subject may make an error and still possess that content. Moreover, this latter component trumps the dispositional account in the following sense. The component that allows for error provides the correct attribution of intentional content in all cases of conflict with the dispositional account. Only if the conditions of application for the component which allows for error are not met, does the dispositional account come into its own. Thus, the basic point is that there may be a lexical ordering amongst accounts of the nature of intentional properties such that part of what it is for an intentional property to occur is that the means of representation that possesses it should not also satisfy those conditions that are sufficient for an occurrence of another intentional property of a character specified by an account higher up the order.

This brings us to the final feature of our approach, its strongly anti-sceptical attitude to intentional content, in taking seriously our actual ascriptions of them to isolate the facts mentioned. It may be asked how we can be so certain that there are such intentional contents if we know so little about some of the facts upon which the possession of an intentional property depends.

The answer is not just that we have a practice of ascribing intentional contents with some stability for we must allow that such a stable practice may still be misguided if we lack firm evidence for the existence of intentional contents. The foundation of our anti-sceptical approach is our understanding of sentences of natural language. We take them as having intentional properties, and this is prima facie evidence that they do have such properties. Moreover, both in being able to have the range of thoughts that can be expressed in language, and in taking the utterance of these sentences by others as an expression of their thoughts, we have indirect access to the very facts that must exist for such a practice to be legitimate, however arcane they may be.

We have found that an account of intentional properties in physical terms is still possible. However, the account we have just put forward faces two difficulties. First, at one stage, when we spoke of cognitive contact, an appeal was made to a subject's perceptual awareness of the item represented. If the argument of the previous chapter is correct, and, such an appeal is ineliminable, then the account of intentional properties would be partly non-physical after all. This

would be especially problematic if perceptual awareness was thought to involve a causal relationship between occurrences of phenomenal properties and the item of which the subject was aware, for the principle that the physical world is causally closed, that we have endorsed, would rule such a relationship out. One might be faced with an argument for the impossibility of some intentional properties as well as an argument for epiphenomenalism.

The second difficulty that the account we have endorsed raises, concerns what we will call the second challenge of epiphenomenalism. It is not alone in this. As we have already noted, most of the accounts considered in this chapter imply that some intentional properties are environment-dependent and thereby face exactly the same challenge. It is to this that we shall now turn.

References

1. c.f. Brian Loar (1987), 'Subjective Intentionality' Philosophical Topics 15
2. c.f. J. Fodor (1975), The Language of Thought (Chs. 1 and 2), (1985), 'Fodor's Guide to Mental Representation: the intelligent auntie's vade-mecum' Mind 94, J. Fodor (1987), Psychosemantics (Appendix); C. Peacocke (1983), Sense and Content (Ch.8); A. Clark (1989), Microcognition (Chs. 7-10); D.C. Dennett (1978) 'Brainwriting and Mind Reading' and 'A Care for the Common Code' in his (1979), Brainstorms).
3. c.f. W.V. Quine (1960), Word and Object (Ch.2); M.A.E. Dummett (1976), 'What is a Theory of Meaning II' (in G. Evans and J. McDowell, eds., Truth and Meaning).
4. c.f. N. Chomsky (1969), 'Quine's Empirical Assumptions' (D. Davidson and J. Hintikka, eds., Words and Objections; E. Craig (1982), 'Meaning, Use and Privacy' Mind 91; J. McDowell (1981), 'Anti-

- Realism and the Epistemology of Understanding' (H. Paret and J. Bouveresse, eds., Meaning and Understanding).
5. c.f. S. Kripke (1984), Wittgenstein, On Rules and Private Language (Chs.1 and 2).
 6. For example, see S. Blackburn (1984), 'The Individual Strikes Back' Synthese 58.
 7. N. Goodman (1976), Languages of Art (Ch.1)
 8. e.g., J. Fodor (1984), 'Semantics, Wisconsin Style', Synthese 59.
 9. J. Fodor (1987), Psychosemantics (Ch.4)
 10. F. Dretske (1981), Knowledge and the Flow of Information (Chs.7-9).
 11. Dretske takes the account to apply only to what he calls 'simple' cases of concept acquisition. Since his example is the concept of a robin presumably or later discussion of the concept of a dog is reasonable case with which to test the theory.
 12. See J. Fodor (1984), 'Semantics, Wisconsin Style', Synthese 59, (1987), Psychosemantics (Ch.4), (1990), A Theory of Content (Chs. 3 and 4).
 13. J. Fodor (1987), J. Fodor (1990), see previous footnote. In the latter, Fodor adds a condition appealing to a means of representation's actual causal history which need not concern us in the criticism that follows.
 14. J. Fodor (1990), p.107-8
 15. I have simplified a little, and put it in the terminology of the present chapter, see F. Dretske (1986), 'Misrepresentation' (R. Bogdan, ed., Belief), (1988), Explaining Behaviour (Chs. 3-4).
 16. R. Millikan (1984), Language, Thought and Other Biological Categories; D. Papineau (1989), Reality and Representation.
 17. c.f. J. Fodor (1990), for a more detailed discussion.
 18. G. Evans (1982), The Varieties of Reference (Ch.9 ss 4, and 5); T. Burge (1979), 'Individualism and the Mental' (P. French, T. Uehling and H. Wettstein, eds., Midwest Studies in Philosophy Vol. IV), (1982), 'Other Bodies' (A Woodfield, ed., Thought and Object), (1982), 'Two Thought Experiments Reviewed', Notre Dame Journal of Formal Logic Vol. 23, (1986), 'Individualism and Psychology' Philosophical Review 95, (1986), 'Intellectual Norms and Foundations of Mind' Journal of Philosophy 83.
 19. Such mental events are sometimes said to have 'broad' 'object dependent' content. The doctrine that asserts that they exist has been called 'anti-individualism' or 'externalism'. There are weak versions of this thesis that we shall not discuss c.f. C. McGinn (1989), Mental Content (Ch.1). Most of our conclusions apply to such versions too.

Chapter 8

The second challenge of epiphenomenalism

The character of intentional properties has often been taken to imply that occurrences of these properties are causally epiphenomenal. Various reasons have been offered for this conclusion. Our argument in Chapter 3 suggests that one could not arrive at it just by showing that intentional properties supervene on other properties, nor by establishing that intentional properties are macro-properties, nor, even, by demonstrating that they are functionally defined properties.¹ However, there is a rather more compelling argument for the claim that if intentional properties are environment-dependent then they are epiphenomenal. It runs as follows:²

- (1) All entities proposed by psychology to be explanatory of the range of phenomena with which it is concerned may be parts of a person independent of the way the person's environment is.
- (2) Some mental events are dependent upon the environment for their existence.
- (3) The mental events mentioned in (2) will not be

cited in any psychological explanation, (from premises (1) and (2)).

- (4) If mental events have any explanatory role to play, then they will be cited in psychological explanations.

Therefore,

- (5) The mental events mentioned in (2) do not have any explanatory role, and are, hence, epiphenomenal.

Our present discussion will be limited to the first premiss, the others, premises (2) and (4), will be taken to be intuitively plausible.

To assess the first premiss, we must make explicit an assumption that arguably lies behind it, namely that:

Psychological explanation is only causal explanation. Although the assumption concerns all psychological explanation, in fact, all that will be relevant to the discussion that follows is that the psychological explanation of behaviour is just causal explanation. The limitation of our discussion to behaviour is for ease of presentation alone. In the subsequent two chapters we will investigate whether the approach to psychological explanation just mentioned is correct. For the moment we will consider whether, if psychological explanation is just causal explanation, environment-dependent mental events have no explanatory role.

First, let us consider the character of those mental events with intentional content that are environment-dependent, for, arguably, only then will we understand the force of the argument. It appears that each such environment-dependent mental event is either composed from, or supervenes upon, two components, a 'head' component and a 'relational', or environment-dependent, component. In some cases these become fused, as we shall see.

To understand the character of the relational component one must consider how it is that a mental event may be dependent upon the environment for its existence. There seem to be three ways. First, we may suppose that the relevant items in the environment are literally part of the mental event. The dependency adverted to would thus be subsumed under the normal dependency of whole upon part. Second, we may suppose that the mental event is composed from a relation having, as one of its relata, an item in the environment. The dependency would then be a matter of the dependency of an occurrence of a relation upon its relata. Third, we may hold that the mental event is composed from a relational property, one which holds because the relation holds. An example of a relational property is that of being two feet to the left of a pile of books. Something possesses that relational property if it is the first of the relata in the relation expressed by

the predicate '- is two feet to the left of -', with the second relata being a pile of books. It is, then, dependent upon whatever the relation is dependent. Arguably, those who wish to assert that environment-dependent mental events supervene upon, but are not composed from both a head and a relational component, will allow that these mental events are composed from a 'relational' component alone of the sort we have just identified.

It is inescapable that there is some head component closely connected to the occurrence of environment-dependent mental events, however esoteric this head component may be taken to be. It is only by supposing that there is such a component that we can understand how a subject is having environment-dependent mental events in one of the ways indicated. If the mental events in question are environment-dependent because they are composed from items in the environment, or a relation with these items as relata, then there must be something that makes these particular items, or relation, and not others, part of a subject's mental life. It is their connection with a head component of the subject which does it. On the other hand, if the mental events are environment-dependent because they are occurrences of relational properties, and thereby supervene upon the elements mentioned a moment ago, then there must be a head

component for the very occurrence of that relational property.

We shall not try to develop our understanding of the nature of environment-dependent mental events further, by assessing the plausibility of the various accounts of the dependency on offer. It is enough that we have established that there is going to be a head component no matter which account is adopted.

The doubt about the explanatory role of environment-dependent mental events arises because it is thought that the environment-dependent aspects of these events lack causal efficacy. In which case, two putatively distinct types of mental events may have the same causal powers if at least one of them is individuated in part by environment-dependent features. But, if psychological explanation is just causal explanation, we should individuate its explanatory entities by their causal powers. Since environment-dependent mental events are not so individuated, they should not appear in psychological explanations.

The challenge to the causal efficacy of the environment-dependent aspect of mental events has been thought to arise in two ways. First, there is the intuition that what causes behaviour must be inside the head. Second, there are intuitions concerning what contentful mental event we would be inclined to say was

causally explanatory of a piece of behaviour.³ It is important to keep these separate. The second line, if distinguished from the first, does not establish the conclusion at which it aims. Roughly speaking, it relies upon the intuition that phenomenally identical environment-dependent mental events do not give rise to different behaviour. Even if this is true, it does not follow that the environment-dependent aspect has no distinctive causal efficacy because we have not established, and it is indeed difficult to establish, that all phenomenal properties are environment-independent. Consequently, we shall focus on the first line of argument.

We can establish that the environment-dependent components have no causal influence upon the causal relations between head components once we remember that each of the arguments that try to demonstrate some mental events are environment-dependent presumes that two subjects may be alike in all other, 'internal', respects and yet have different mental events of the environment-dependent sort. Here is the argument, informally expressed.

- (1) If the environment-dependent aspects have an affect upon the causal relations between head components, then different such aspects give rise to different causal relations.

- (2) If different environment-dependent aspects give rise to different causal relations between the head components, then two subjects may not have different environment-dependent mental events without being different in other, 'internal', respects.
- (3) Two subjects may be alike in all 'internal' respects yet possess different environment-dependent mental events.

Therefore:

- (4) It is not the case that the environment-dependent aspects have an affect upon the causal relations between head components.

It is the second premiss which is crucial. The existence of different causal relations between head components implies that they have a different, internal, causal role. Thus, we have an internal difference to correspond to every causally significant environment-dependent aspect. Even if such differences were overlooked, we would still have to consider their explanation. Either the difference is a consequence of features of the head component, or it is the result of action at a distance, by the environment. The latter has been rejected in this case.⁴ But, to adopt the alternative is to admit that there are corresponding internal differences after all.

We have not yet reached the conclusion that

environment-dependent aspect of these mental events lack causal efficacy. All that we have found so far is that only the head components are causally efficacious with regard to the occurrence of other head components. However, the conclusion is not far away. If it was true that:

- (5) Only the head components that compose environment-dependent mental events or upon which environment-dependent mental events supervene, have, in conjunction, distinctive causal consequences upon the behaviour of the subject whose head they partly constitute,

then it would follow that:

- (6) The environment-dependent aspects of mental events are causally epiphenomenal with regard to behaviour.

If environment-dependent mental events were taken to be just the occurrence of relational properties, they would be causally epiphenomenal. If not, they would have causal efficacy as a result of their head component alone. The rest of this section will be focused on the issues raised in connection with premiss (5).

It has been claimed that there are behavioural differences that correspond to differences in the environment-dependent aspects of mental events.⁵ The descriptions of a subject's behaviour often makes

reference to what has been successfully achieved, or what the subject is trying to achieve, and this will depend, in part, upon how the environment is at some point. Assuming that the descriptions mentioned reveal the nature of behaviour, this would make behaviour environment-dependent. For example, one may only describe someone as turning on the hot water tap, or as trying to turn it on, if there is a hot water tap there, or, in the second case, if the subject has come across a hot water tap before to think of trying to turn it on.

Now, the claims just made are contentious. It may be possible to describe behaviour in a way which is legitimate and not dependent upon the environment being a certain way. However, let us work upon the supposition that behaviour is to be described in environment-dependent fashion. One can still hope to show that, contrary to what has been asserted,⁶ the environment-dependent aspects of mental events are not causally efficacious with regard to such behaviour.

Jerry Fodor⁷ has provided the main argument against the claim that differences in the environment-dependent part of mental events cause differences in behaviour environment-dependently described. He notes, first, that not just any relation between an item in the environment and some entity can confer a causal power upon that entity. To use his example, each event in the universe

can be identified as an event in a universe in which a particular coin is facing up heads or tails, hereafter head-type and tail-type events respectively. Now consider each cause and effect in the universe. Since we have chosen to individuate events in the universe by, in part, their relations to that coin, these cause and effect pairs will be different depending upon whether the coin is facing up heads or tails. From which it follows that a cause in the universe in which the coin is turned up heads will bring about something different from what we might be tempted to call the same cause in a universe in which the coin is turned up tails. Nevertheless, says Fodor, we would still not want to count as causally efficacious the relational property which depends upon which side of the coin is face up.

It is hard not to share Fodor's intuition with regard to this type of case. But, if we are going to use the intuition to undermine the putative causal efficacy of the environment-dependent aspect of mental events we need to explain why we think the 'coin-typed' events do not have distinct causal powers. Fodor's own suggestion is that if the relation to the side of the coin facing upwards was part of the causal powers of the objects in the universe, then there must be a causal mechanism or fundamental law which showed how the causal powers depended upon it. He then asserts that we have good reason to suppose there is

no such causal mechanism or fundamental law in the case of the 'coin-typed' events or in the case of environment-dependent mental events and their causation of behaviour.

Martin Davies⁸ points out that the demand for a causal mechanism or fundamental law is potentially question-begging since those who hold that the environment-dependent aspects of mental events are causally efficacious suppose that the causal powers of an entity may be metaphysically dependent, or, at least, not causally dependent, upon items in the environment in much the same way as the occurrence of the entity is. So, we cannot appeal to Fodor's suggestion as to what rules out the coin-typed events having distinctive causal powers, unless we want to merely assert, and not argue for the claim, that the environment-dependent aspects of mental events lack causal efficacy. Also, in the case of environment-dependent mental events, the requirement that there is some sort of causal mechanism or law rests upon the assumption that the psychological efficacy of these mental events must be located wholly within the subject's skin, so to speak. But this is the very matter which is at issue and cannot just be assumed.⁹

The real reason why we should not allow that the coin-typed events have distinct causal powers is that to do so would be to infringe the following principle.

A property of an entity, or the occurrence of a property, should not be considered to have distinct causal consequences if its possession, or occurrence, is not a necessary condition of those very consequences occurring for a specific collection of causes.

Consider first the coin universe. Suppose there is a head-type event caused by another head-type event. The cause is, of course, a necessary condition of this effect, but, is it a necessary condition of that component of the effect which makes it a head-type event? Surely the answer is no. This component of the effect would have occurred if the coin was only turned up heads after the time of occurrence of the cause, and, would not have occurred had the coin not been turned up heads at the time of occurrence of the effect, even if the causes had been head-type events. It seems reasonable to take this as evidence that the occurrence of head-type causes was not necessary for the production of head-type effects. Otherwise, we might have to allow there was an overdetermination of that component of the effect.

Of course, it does not follow that the side of the coin is a cause of an event being either a head-type or a tail-type event. It is just that one should not count something as a distinct causal consequence of something else if the putative causal consequence's occurrence

varies with regard to some other factor and not our candidate cause. How one accounts for the variation is another matter.

Consider, now, environment-dependent mental events. We have noted that they are not always dependent upon the present environment for their existence. In contrast, the first type of environment-dependent behaviour we identified is dependent upon items in the immediate environment. So, the remarks that we made in the case of the coin universes have straightforward application. If I am picking up a bar of gold, there must be gold present to be picked up. However, the presence of gold is in no way causally determined by the prior occurrence of, arguably, environment-dependent mental events such as desiring a bar of gold and believing that there is a bar of gold before one.

Unfortunately, the other type of environment-dependent behaviour cannot be dealt with in the same manner. What someone tries to do depends upon their prior mental events, and only if the latter are dependent upon an environment is the behaviour also dependent upon it. For example, someone may be trying to pick up some gold, even if no gold is present, because they believe it to be present. But, arguably, it is not possible that he or she could try to do this if he or she did not have a belief of this kind concerning gold, and such beliefs

could only be had if the subject had at some point come across gold in his or her environment. A similar line of thought can be applied to verbal behaviour such as asking for some water.¹⁰ In each case, the point of the ascription of such behaviour seems to be to capture the subject's view of what he or she is doing specifically in terms of the beliefs, desires, and resultant intentions that he or she has. Consequently, it seems that one genuinely could not say that a piece of this sort of behaviour occurred in the absence of the relevant type of mental events, including environment-dependent mental events, and, thus, the environment-dependent aspects of these mental events can be cited to explain why the relevant behaviour occurred.

Fodor, in effect, tries to deal with the problem by claiming that it is a conceptual truth that differences in the environment-dependent aspects of mental events are required for differences in the environment-dependent aspects of trying to do something.¹¹ He argues that since conceptual connections rule out causal connections, it has not been shown that the environment-dependent aspects have distinct causal consequences. However, these are controversial claims to make and it would be better if we could avoid them. Some have been wont to claim that all causal statements are conceptual truths because, to put it crudely, all properties are powers.¹² Others might not

see the statement of the variation as a conceptual truth. It is certainly the case that the existence of a conceptual relationship between the description of two things is compatible with a causal relationship between them.¹³

Instead, we should adopt a somewhat more cautious approach. Intuitively, we can agree that environment-dependent mental events are necessary for certain sorts of tryings. We may also have grounds for supposing that the necessity is in some sense 'conceptual'. Our concept of this sort of behaviour seems primarily the concept of a type of behaviour entirely characterised in terms of certain of the mental events which led up to it. However, we need not commit ourselves to this thesis to throw into question the putative causal efficacy of the environment-dependent aspects of mental events in this regard. Instead, we can note that because of the possible conceptual connection the dependency that we have recognised between the environment-dependent aspects of the mental events and behaviour need not have a causal origin. There is another explanation of it. Thus, the justification for believing in a causal relationship is commensurately weaker. One might then cite one's intuition that the causation of behaviour is localised beneath the subject's skin, so to speak, as evidence against the conceptual connection occurring along-side a

causal one in the case in which we are interested.

Although, there is nothing like a demonstration that the environment-dependent aspects of mental events lack causal efficacy to show for our discussion, one should be, at least, wary of the claim that there is no such threat. So, it is worth considering what we should say if the environment-dependent aspects are causally epiphenomenal.

All things being equal, it would seem to follow that if one is committed to psychological explanation being merely causal, one should not individuate its entities by reference to their environment-dependent types. However, it would not follow from this that we should not individuate mental events by the full range of intentional contents, for individuating mental events in this way is not to use an 'environment-dependent' typology. For every type of intentional content, there are both environment-dependent and environment-independent occurrences.¹⁴ For example, a subject may have environment-dependent intentional content involving water if he or she does not have the wherewithal to be able to identify water independently of the environment but still should be ascribed such content for the sort of reason we canvassed at the end of the previous chapter. On the other hand, if the subject is able to identify water, and the grounds of this ability is not dependent upon the way the environment is, then that subject may be ascribed an environment-

independent intentional content involving water. If one knew that the referent of 'water' was H₂O, then, arguably, one could have beliefs about water without ever having come across the stuff.

Individuation by intentional contents is, in fact, an individuation by causal powers. All distinct intentional contents have distinct causal consequences within those subjects who are able to recognise their distinctness. Individuation by causal powers does not require that an entity so individuated should bring a distinctive contribution to every context in which it occurs. So, the existence of environment-dependent occurrences of intentional contents does not undermine the practice of categorising by intentional content. Bearing these points in mind, the problem of epiphenomenalism we face is not about intentional contents in general, but about a particular use of them to explain behaviour. The question is, why should we explain a subject's behaviour by an environment-dependent mental event, when there is some environment-independent property, whose occurrence we might cite, which in fact did the causal work? We will try to answer this question in the next chapter

References

1. Contrary to what is said in N. Block (1987), 'Can the mind change the world?' (in G. Boolos (1990), ed., Meaning and Method).
2. S. Stich (1978), 'Autonomous Psychology and the Belief-Desire Thesis' The Monist 61.
3. J. Fodor (1980), 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology' The Behavioural and Brain Sciences Vol.3; G. Segal (1989), 'The Return of the Individual' Mind 98.
4. e.g., T. Burge (1989), 'Individuation and Causation in Psychology', Pacific Philosophical Quarterly 20.
5. J. Hornsby (1986), 'Physicalist Thinking and Conceptions of Behaviour' (P. Pettit and J. McDowell, eds., Subject, Thought and Context.); G. Evans (1982), The Varieties of Reference (p.200-204).
6. c.f. J. Hornsby (1986), T. Burge (1989).
7. J. Fodor (1987), Psychosemantics (Ch.2), (1991), 'A Modal Argument for Narrow Content' Journal of Philosophy 88.
8. I have abstracted a little from the other issues discussed to highlight the point I wished to raise, see M. Davies (1986), 'Individualism and Supervenience', Proceedings of the Aristotelian Society, Supplementary Volume 60.
9. Pointed out in both M. Davies (1986) and T. Burge (1989).
10. The issue is raised, more or less, in T. Burge (1989), 'Individuation and Causation in Psychology' Pacific Philosophical Quarterly 70 and discussed in J. Fodor (1991), 'A Modal Argument for Narrow Content' Journal of Philosophy 88.
11. J. Fodor (1991).
12. S. Shoemaker (1980), 'Causality and Properties' P van Inwagen (ed.), Time and Cause.
13. c.f. A.J. Ayer (1973), The Central Questions of Philosophy (p.141).
14. A possible, but in my view unlikely, exception is demonstratives and indexicals. See, for instance, J.R. Searle (1983), Intentionality (Ch.8); S. Schiffer (1978), 'The Basis of Reference', Erkenntnis 13.

Chapter 9

On the possibility of a species of explanation based on norms

In this chapter we shall assume that the arguments for the existence of mental events with environment-dependent intentional contents are sound.¹ In the light of the argument of the preceding chapter, this raises the question of whether a more accurate causal explanation of a subject's behaviour should eschew mention of mental events involving environment-dependent instances of intentional contents in favour of what was actually causally active in the circumstances, namely the head components of such events. We shall assume that the answer to this question is yes, and that the only reason for then resisting alteration in our explanatory practices is that we believe that mention of these events plays a non-causal explanatory role. This chapter will be concerned with the attempt to develop such a role in terms of norms.

(a) What is an explanation based upon norms?

The basic idea behind an explanation based upon norms, hereafter, a 'norm-based explanation', is that something is nominated as a norm, for instance, truth

reproduction or rationality, and it is suggested that a certain system of entities should be understood as either trying to act in conformity with, or, acting in conformity with, or, aiming at, or, moving towards, that norm, ('aiming at' or 'trying to act in conformity with' if the system is an agent 'moving towards' or 'acting in conformity with' otherwise). Given that it is appropriate to see a system in this way, the following question becomes relevant for it, with regard to the nominated end: 'How is the system S trying to act in conformity with, acting in conformity with, aimed at, or moving towards Norm N?' That which is to be explained, the explanandum, is, in other words, the particular orientation of the system. The question concerns not the cause of the orientation, but rather what constitutes this orientation. It is for this reason that the explanatory question is better formulated as a 'how question' rather than a 'why question'. The explanans, that which does the explaining, are the components of the system, and the relations between the components, that together constitute the fact of the system being orientated in the way described. One identifies the components of the system by ascribing to each a specific contribution to the norm which is closest to the actual contribution they make, through their relations, largely causal relations, to other components of the system.

If the system were designed by God to achieve the norm nominated, then actual and ascribed contributions would coincide. However, in a godless, or at any rate, imperfect world we should not expect this. When a system is not perfectly acting in conformity with, or moving towards some goal, it may seem questionable whether we should try to see the system in these terms, and even more questionable whether, if we did, we should count the resulting 'explanation' as an explanation. The second matter concerning explanation is both important and apt to be trivialised. We shall return to it in section (d). We shall try to deal with the first matter now.

What we require is some fact about the system which legitimates the adoption of a norm-based perspective on those occasions when the system is not perfectly orientated towards a particular norm. One such fact is that the system has been designed by a thinking creature. Our justification, in this case would be that we wanted to see how the designer envisaged that, by putting together the various parts of the system, the system would have a particular orientation. At the same time, we would, by considering the intention of the designer, see in what way the system failed to measure up to requirements.

A second fact about a system which would justify our adoption of the norm-based perspective is that the system sees itself as aimed at or trying to act in conformity

with the norm in question. Obviously, this form of justification would be limited to self-interpreting systems, and, thereby, arguably to minds. In this case, the adoption of the norm-based perspective would enable us to understand how the system saw itself.

Another range of facts that would provide us with a reason for adopting the norm-based perspective is that the system either desires to be moving towards, or to act in conformity with, the norm, or, has some analogue of these desires. In this case, the type of understanding we would obtain would be similar to that we discussed with respect to the system that was designed.

Finally, if there was a system which behaved pretty much as if it was moving towards, or, acting in conformity with, the norm, then one might wish to adopt the norm-based perspective to determine the extent to which, and how, this is so. The main motivation in this case, would be that the norm was of particular interest to us, or provided a particularly neat way of describing the activity of the system in general.

(b) Can the environment-dependent character of some mental events be given an explanatory role in terms of the norm of truth?

The idea that we shall consider, in this section, is that the environment-dependent aspects of mental events

have a norm-based explanatory role. To begin with, we need to identify the relevant norm at which minds are aimed. The suggestion that we shall take up here is that, crudely speaking, the norm is truth. We may put this slightly more precisely. Consider a list of propositions that collectively exhaust what may be truly said of the world. One of the aims which minds have is to believe the propositions on that list and not to believe propositions that are not from that list. It is in this sense that minds are aimed at the truth.

We may justify the claim that minds are aimed at the truth by referring to the four types of facts we said indicated that a norm-based explanation is appropriate. In particular, we may argue that, at least, human minds are aimed at the truth because that is how it seems to them, that is how they desire themselves to be, and that is roughly how they behave, with regard to a norm in which we are interested. Now, of course, these claims may be questioned. However, we shall presume that they are true for the discussion that follows. All that we are ultimately committed to is the claim that it would not be correct to ascribe environment-dependent mental events to anybody for whom one of these facts did not hold.

What we must now do is explain how environment-dependent mental events have an explanatory role with the type of norm-based explanation we have been discussing. A

preliminary point to make is that the partial individuation of some mental events by their intentional contents is appropriate for the norm-based explanation envisaged. Propositional intentional contents have truth conditions. So, when the intentional contents are coupled with another component of a subject's mind to form a belief, we have that type of entity which, by its presence, indicates the extent to which the subject has attained, or failed to attain, the norm of truth, as we characterised it. However, although intentional contents have a nature which makes them suitable to occur in this sort of explanation, one might still wonder whether the ascription of environment-dependent intentional content provides any explanatory insight into behaviour, especially when, as we have already recognised, they are not causally active, although attributed to something which is.

We shall try to delineate the explanatory role that environment-dependent mental events have by considering, in some detail, an example. Suppose that Jack is presently sipping a brandy he has been given by his hostess. Suppose, further, that to his relatively unsophisticated palate the taste of all brandies are much the same and that they all look the same to him. At a suitable moment, he says 'This is a fine cognac' and pours himself another glass. It is natural to suppose that the explanation of

his behaviour is that he desired some more cognac.

As a matter of fact, Jack's hostess has provided him with an armagnac. Does it follow that we have no explanation of Jack's behaviour, which was, in fact, the pouring of armagnac into his glass? One is inclined to say no. The desire that we ascribe to him was a consequence of the desire for a certain kind of taste, plus the belief that brandy tastes like that, and that all brandy is cognac. So, we are tempted to say that what really motivated his behaviour was the desire for a certain kind of taste and the belief that if he poured himself another glass, then he would be more likely to obtain it. However, it is at this point that one must be careful. We can admit that the belief and desire just mentioned are one explanation of the behaviour, perhaps even the most important. Nevertheless, they are not the only explanation. We can explain the very same behaviour in terms of the desire for cognac, the belief that that liquid is cognac and the belief that if he poured himself another glass, then he would be more likely to obtain cognac. The difference between the two explanations is that the second contains more of what we intuitively would count as environment-dependent mental events than the first, by mentioning beliefs and desires concerning cognac, something that Jack does not have the wherewithal to identify independently of the environment.

Is one of the explanations preferable to the other? We shall assume that both explanations are relevant to the behaviour, in the sense that the head components of the mental events identified are causally efficacious of the behaviour, but that the mental events cited in each explanation, on pain of overdetermination, are not part of the same collection of causes.² If our only interest is in causal explanation, arguably, the first is preferable. Although it does not distinguish all of the mental events by features which are actually causally efficacious in the circumstances, it mentions more efficacious elements than the second explanation does. However, if we broaden our understanding of explanation we can see that the second provides us with further insight into the behaviour of Jack. It reveals to us how the behaviour relates to his other beliefs about, and attitudes towards, the world, and thereby, integrates his behaviour into his attempts to find out the truth and justify what he believes. It is not that when we have put forward the first explanation there is nothing more to understand about the causal antecedents of his behaviour.

Clearly if we are to allow that the second explanation provides insight, then we are, in effect, agreeing that there is a non-causal explanatory role. We may dub this an 'inferential role'. The ascription of each mental event with intentional content fills in a gap

in the pattern of deliberation that eventually gives rise to a piece of behaviour, and each gap so filled is an inferential gap in the sense that it is defined by reference to deductive and 'inductive' logic. For example, the gap between desiring cognac, and pouring armagnac is partly filled by supposing that the subject believes that that liquid which he is to pour is cognac.

The ascription of environment-dependent intentional contents also provide us with insight into the causal antecedents of a piece of behaviour by revealing some cognitive dysfunctions that might have been amongst the antecedents. For example, if we ascribe to Jack a belief that he is drinking fine cognac, then given the fact he is drinking armagnac we need an explanation of why he believes what he does, the sort of explanation we pointed to by describing his crude tastes and his beliefs concerning brandy. There would have been no pressure to provide this very explanation if the environment-dependent intentional content expressed by, say, 'This is a fine cognac I am drinking' had not been ascribed.

Now, it might be argued that we cannot justify our provision of an explanatory role for environment-dependent intentional contents because we have merely presumed that such intentional contents are ascribable. However, our aim was not to defend their ascription but to show that if they were ascribable, then they would have an explanatory

utility. So, the overall force of the argument is that given that we intuitively think that such intentional contents are ascribable, then there is no need to revise our intuitions because these intentional contents have no explanatory worth.

Our discussion of the example reveals the relationship between what we have called a 'norm-based' explanation of behaviour and an 'intentional' or 'rational' explanation of it, where the latter is concerned with the reasons for a piece of behaviour and identifies them by ascribing to the relevant subject beliefs and desires. Plausibly, intentional or rational explanation is a species of norm-based explanation appealing to the norm of rationality.³ However, it might be felt that intentional explanation need not mention environment-dependent mental events for the reasons we have already mentioned. So, the provision of a norm-based explanation of behaviour which mentions the latter type of mental events involves one in extra commitments. As some philosophers have noted, the resistance to the use of environment-dependent mental events may make certain behaviour inexplicable by the citation solely of mental events with intentional content.⁴ For instance, arguably, if my picking up of a particular glass is to receive an intentional explanation, one must mention a belief with an intentional content specifically about that glass. Since

it is unlikely that what is phenomenologically before my mind will be sufficient to single out that glass, it is tempting to suppose that if I do have the relevant belief it will be environment-dependent. Thus, if we are to fully explain that behaviour by noting only mental events with intentional content and not by having to provide an independent specification of the context in which the agent is about to act, an environment-dependent mental event will be required.

The basic idea of norm-based explanation of behaviour, then, is as follows. A necessary condition upon the explanation of behaviour is that we cite the behaviour's causes. In so doing, we provide the only answer possible to the question, why did the behaviour occur? Also, in stating what caused the behaviour we indicate the components of the subject relevant to the explanation of behaviour thus limiting the norm-based explanation that may be given. If the subject satisfies one of the conditions that makes a norm-based explanation appropriate, we may then consider the additional question, how is that behaviour the causal consequence of states of a subject who is aimed at the truth? The correct answer will cite the intentional contents of the causally relevant mental events and take the form we have already indicated. The norm-based explanation we have just identified is also an integral part of a rational

explanation of the behaviour. The fact that it is legitimate to ascribe to the subject environment-dependent intentional contents implies that the type of beliefs and desires cited in the rational explanation of his or her behaviour may themselves be environment-dependent. It is in these two ways that environment-dependent mental events have a role to play in the explanation of a piece of behaviour over and above the causal contribution they make.

(c) A comparison of our proposal with other models of explanation

In this section, we shall contrast our approach with others that have been adopted. Each of the accounts below try to supply an explanatory role for the environment-dependent aspect of mental events and it is their success in this, that will decide whether or not they should be accepted. However, just as important as an assessment of their capacity to do this, is the potential illumination they supply by highlighting some of the features of our own account.

One proposal that has been put forward, by Frank Jackson and Philip Pettit,⁵ is that when we cite environment-dependent mental events to causally explain behaviour we put forward a species of 'program' explanation. Jackson and Pettit reject the claim that if

e causally explains f, then e is causally efficacious of f. Instead, they suggest that if there is a range of properties each of whose occurrence may be causally efficacious of a certain effect of type f, and one of whose occurrence has in fact been causally efficacious of an effect of type f, then a description true of that range of properties alone can be used to causally explain the effect in question. They argue that when we refer to environment-dependent mental events in the explanation of behaviour we have a case of 'program' explanation. Such mental events have an explanatory role because the ascription of an environment-dependent mental event in effect picks out pairs of head components and environment components each of which would be causally efficacious with regard to a certain piece of behaviour, and one pair of which is, in the circumstances.

In contrast to our own proposal, their's sticks within the framework of causal explanation. However, Jackson and Pettit do not oppose the conclusion we reached in the preceding chapter. They are prepared to agree that in some sense environment-dependent mental events are not really causally efficacious. So one might take their distinction between 'causally efficacious of' and 'causally explain' as a distinction between explanatory role and explanatory utility. Ascriptions of environment-dependent mental events provide us with handy ways of

grouping causally efficacious entities, and hence have explanatory utility, but they do not have any special contribution to make to the final understanding of what is going on in the world, they have no explanatory role. Our proposal was more ambitious in that it attempted to provide environment-dependent mental events with the latter. Thus, all things being equal, since it provides more, it should be adopted.

Fred Dretske⁶ also believes that intentional contents, including environment-dependent occurrences of them, have a role to play in the causal explanation of behaviour. He believes that the difference between a piece of behaviour, such as moving one's arm, and a movement, such as having one's arm moved, is that all behaviour is the causal consequence of some means of representation with a particular intentional property. As we saw earlier, according to Dretske, means of representation only cause behaviour because they have been selected for this role as a result of their correlation with an item in the environment and it is in this selection that he sees the causal activity of intentional properties.

We have already seen that there are problems with Dretske's account of intentional properties, so we should be cautious of adopting his approach to their explanatory role. Given that our previous criticisms are correct, it

would seem that even if Dretske is right in what he says about the causal correlation between a means of representation and an item in the environment being itself causally efficacious, he would not have demonstrated the causal efficacy of intentional properties. However, it is not clear that a potential means of representation's causal correlation with an item in the environment is itself causally efficacious. A similar line of reasoning to that contained in Chapter 8 would, more than likely, threaten it by holding that all the causal activity takes place as a consequence of the causal relata, and of the effects of the relata, and not because of the correlation. So, at best, it might be thought that Dretske's account does not explicitly address the worries to which our account is addressed.

A third possibility is that intentional contents generally have a 'functional' explanatory role.⁷ Once more, the basic idea is that intentional properties are those properties that are possessed by means of representation because the latter have a certain function. We may then envisage that such properties will be cited in the following two types of explanation. First, by their ascription to means of representation, we understand how the system is meant to work, that is, we may explain the occurrence of other things by noting that it was the function of means of representation to bring them about.

For example, we may functionally explain why my desire for an apple causes me to eat an apple by saying that it is the function of such a desire to have that behaviour as a consequence.

Second, one may functionally explain the occurrence of things causally prior to the means of representation with the functions indicated, by saying that they occurred because they were disposed to give rise to the means of representation in question.⁸ For example, suppose one asks for a functional explanation of why a subject examined the contents of his or her desk drawer very carefully. One might give as an answer: because by so doing will he or she find out whether there is something in it he or she wants. The searching activity is functionally explained by its likely result, finding out something and the latter, it might be thought, is constituted from, in part, the occurrence of means of representation with intentional properties.

Although there are many similarities between the approach just described and our own, there are also differences of emphasis that need to be noted. First, we are more liberal in the number of facts we said would licence the provision of a norm-based explanation, than is common for functional explanations. Usually the latter are only considered legitimate if either a system has a feedback mechanism or has been designed. Second, nothing

in what we have said so far commits us to holding that intentional properties are functional properties, or that, they have a function.

(d) On the status of norm-based explanation

Up until now, we have presumed that the categorisation of norm-based explanation as explanation rather than just description, say, is uncontroversial. However, this is unlikely to be so. A threat to the status of norm-based explanation may come from, at least, two areas. One group of antagonists hold that causal explanation is the only legitimate form of explanation. A second group may be indifferent as to whether norm-based explanation is explanation but claim that it is not psychological explanation. We shall consider the objections that the latter may bring in the subsequent chapter. For the moment, we will focus on those who claim that the only legitimate form of explanation is causal.

One objection that those who favour causal explanation may bring arises from the following reasoning.⁹ The only request for explanation that we should seek to answer is the following: 'Why did something occur?' The only answer that may be given to such a question is that since something else, the cause, happened it had to occur. Thus, all explanation is just the explanation of certain events, the effects, by the

citation of others, the causes.

As we are committed to rejecting the claim that the only question we should seek to answer is 'Why did something occur?', we could just say that the described view is inadequately motivated since no reason is given as to why the explanatory questions that we favour are illegitimate, and the question they favour legitimate. However, this is not a very satisfactory reply since our aim was to justify our categorisation of norm-based explanation as explanation and not just to show that nobody had demonstrated that it was unjustified.

In order to do better, we need to consider what it is to justify calling a particular type of explanation 'explanation'. One thought would be that we can somehow appeal to the meaning of the word 'explanation' to determine what should count as an explanation. If this was all we had to do, our task of justifying our categorisation of 'norm-based explanation' as explanation would be reasonably easy. First, if use is any guide to meaning, then the meaning of explanation is rather broader than those who assert that only causal explanation is explanation need it to be. Second, the 'explanations' of behaviour we offered, earlier, intuitively were explanations. Moreover, they were not just explanations because they were causal. The intentional contents provided what, *prima facie*, seemed like another

explanatory dimension.

As a consequence of the points just made, it seems reasonable to suppose that those who exalt causal explanation are putting forward a prescription, as to what we should count as explanation, which is somewhat revisionary. It is an attempt to clean up a loose practice. Such a revisionary posture is, of course, itself in need of some justification. So, we should consider what this might be.

The following hypothesis seems worth considering:

Something is an explanation if it provides understanding of that which it explains.

A causal explanation, it may be said, provides us with an understanding of an event by supplying us with those conditions which were such that, to put it crudely, if they held, then the event either had to happen, or was more likely to happen, if the cause was probabilistic.¹⁰ However, norm-based explanation does not provide us with an understanding of that which it explains. We do not understand how a system works, say, as a result of being provided with that type of explanation.

Unfortunately for those who favour causal explanation, it is just unclear why we cannot be said to understand something about a system by describing its activity by reference to a particular norm. In fact, the four justifications we gave for the application of a norm-

based perspective were formulated in terms of what there was to understand.

At this point, there are two possibilities open to our opponents. First, they may provide some other justification of their revisionary approach. But, it is not obvious that there is an alternative which, on the one hand, is strict enough to validate causal explanation without validating norm-based explanation as explanation, yet, on the other hand, is sufficiently general that it seems to capture a truth about explanation and not just about their particular preference. One is tempted to think that our opponents will fall back upon the assertion that only a certain type of question is a request for explanation, and only a certain type of answer to this question may be given. An approach we have already tried to characterise as unsatisfactory.

The other possibility open to our opponents is to agree that there is a link between explanation and understanding as we suggest, but to claim that the notion of understanding, as it is appealed to in the hypothesis above, is underdeveloped. When we obtain something better, the hypothesis will correctly pronounce that only causal explanation is the genuine article.

We can, of course, agree that our present understanding of understanding is poor. However, it is hard to be confident that, when it is developed, we will

see that the norm-based explanations cited earlier provide us with no understanding of a subject's behaviour that stretches beyond their causal aspects. Once more, one is inclined to think that the sort of dilemma that we mentioned in our discussion of the first possibility will be faced here too.

Thus, our defence of norm-based explanation as explanation amounts to this. The only justification that may be offered for categorising something as an explanation is that it helps us understand the thing it putatively explains. According to this, our substantive justification, norm-based explanation is genuine. Any other putative justification trivialises the whole debate.

During the course of the present chapter, we have tried to indicate a way in which environment-dependent mental events may be seen to have an explanatory role. They would have it, we claimed, by playing a distinctive role in a member of the species of explanation we called norm-based explanation. We then contrasted such explanation with other accounts that have been provided and defended it against possible criticisms of its status. It remains for us to deal with the last source of threat to our rejection of the first premiss of the second

challenge of epiphenomenalism. It may be held that whatever norm-based explanation of the sort indicated is, it is not psychological explanation.

References

1. See references in footnote 15 of Chapter 7 for the particular arguments offered.
2. c.f. D. Davidson (1963), 'Actions, Reasons and Causes', Journal of Philosophy 60.
3. c.f. D. Davidson (1963), reference in footnote 2, and D. Davidson (1970), 'How is Weakness of Will Possible?', (1970), 'Mental Events', (1974), 'Psychology as Philosophy' (all in his, (1980), Essays on Actions and Events).
4. C. Peacocke (1981), 'Demonstrative Thought and Psychological Explanation', Synthese 49.
5. F. Jackson and P. Pettit (1988), 'Functionalism and Broad Content', Mind 97, F. Jackson and P. Pettit (1990), 'Program Explanation, A General Perspective', Analysis 50.
6. F.I. Dretske (1988), 'The Explanatory Role of Content' (D. Merrill and R. Grimm, eds., Contents of Thought), F.I. Dretske (1988), Explaining Behaviour, (1989), 'Reasons and Causes' (J. Tomberlin, ed., Philosophical Perspectives Vol.3).
7. See references in previous chapter to Millikan and Papineau, also C. McGinn (1989), Mental Content, (Ch.2).
8. This appeals to a standard view of functional explanation, e.g., in G. Cohen (1978), Karl Marx's Theory of History. For criticisms of such an account, see C.G. Hempel (1959) 'The Logic of Functional Analysis' in his (1965), Aspects of Scientific Explanation) and T. Honderich (1982), 'Against Teleological Historical Materialism', Inquiry 25. Cohen deals with Hempel's criticisms in the work cited.
9. T. Honderich, in conversation.
10. How one understands this is of course, controversial, c.f. our account of causation in Chapter 2 and, in contrast, C. G. Hempel (1965), 'Aspects of Scientific Explanation' (in his, Aspects of Scientific Explanation); D. Hume (1739), A Treatise of Human Nature (I.iii), (1748), Enquiry Concerning Human Understanding (s4-7).

Chapter 10

On the likelihood of integration of norm-based explanation into psychology

Although it seems that we have some grounds for believing that it is legitimate to cite a mental event with environment-dependent intentional content as part of an explanation of a piece of behaviour, we are far from establishing that such an event should be part of a psychological explanation of it. To complete the reply to the second challenge of epiphenomenalism, we shall have to come to some conclusion about whether the style of norm-based explanation described in the previous chapter can be integrated into scientific psychology. Until then we have the possibility of the following epiphenomenal thesis being true:

The environment-dependent intentional contents of mental events have no role to play in psychological explanation.

And, one might feel, the truth of the latter thesis will be just as disturbing to our intuitive way of looking at the matter as the claim that they have no explanatory role at all, especially since it is probable that there are more environment-dependent instances of intentional

content than environment-independent ones. We would have to accept, in these circumstances, that we were wrong to presume that intentional content had a central role in psychological explanation and thus impute to ourselves a degree of confusion on this subject that we might be unprepared to face.

The approach that we shall adopt must be provisional in character since there is no general agreement as to the type of explanation that psychology should be concerned to provide. We will begin by discussing, only too briefly, that type of explanation currently put forward in the part of psychology most likely to allow that mental events with intentional content have an explanatory role, cognitive psychology or cognitive science (we shall use these terms interchangeably). After we have provided some characterisation of the type of explanation in question, we will attempt to put forward a positive case for counting norm-based explanation as the type of explanation in which scientific psychology should be interested.

1. Explanations as they occur in cognitive science

The type of explanations that cognitive scientists aim to provide, and the methodology that they adopt for obtaining these explanations, has been clearly laid out by David Marr in his book on vision.¹ Marr identifies three levels at which an information processing device, such as

a computer, may be understood. These levels will have direct application to our understanding of mind, since it is the driving intuition behind cognitive science that we shall understand minds by thinking of them as information processors.

The first level Marr calls the Computational level. His idea is that at this level we should identify:

- (a) the goal of a particular, purportedly computational, process, that is, what is being computed;
- (b) the reason why the process has this goal, which is shown by relating it to the overall goal of the system in which it plays a part;
- (c) the mapping of information as input to information as output which is definitive of the achievement of the goal identified.
- (d) the type of information input that is required for (c)
- (e) the general constraints that should govern such a mapping, which constraints in some way individuate the mapping. Here, the idea is, roughly, that just as certain equalities may be used to define the operation of addition - $a+b=b+a$, $(a+b)+c=a+(b+c)$, $a+\text{inverse } a=0$ - so, general constraints upon the mapping of information may define the character of that

mapping.²

Thus, one might say of a particular system, such as the human visual system, that it will have to be able to obtain information about the contours of objects from the stimulation that the visual system receives. In stating that information about contours is what is required, we are stating one of the goals of the human visual system. By providing some characterisation of the type of information contained in the input we determine the task that the computational process must be able to fulfil. The characterisation in the present case may be that the information contained in the input concerns changes in light intensity. The general constraints which are supposed to reveal the character of the mapping are those such as: sharp changes in light intensity reveal edges.

The second level of understanding of information processing systems is what Marr calls the level of Representation and Algorithm. It is concerned with a specific way in which the goals outlined by the computational theory may be achieved. So, the aim is to discover the type of means of representation used. For example, does the information processing system have a language, or is there some other form of means of representation? If there is language, what type of language is it? There is no guarantee that the type of means of representation that mediates the input will be

the same as the type which mediates the output. So, this would be something else which one must discover. Clearly, the type of means of representation that exists will constrain the types of algorithm that may be operative. An algorithm is a collection of rules that maps the relevant inputs to the required outputs. By definition, it has the following features: it maps in all cases that it should, the mapping does not go wrong, it is finite, and it is mechanical.

To bring out the nature of the second level of description we shall contrast it with the third and final level, that of hardware implementation. For Marr, this level involves us in a specification of how a physical object, such as a computer, can have means of representation and manipulate these in accordance with an algorithm. The third level adverts to properties of the computer which are the subject matter of physics and notes how the relations between these properties gives rise to certain patterns of causal activity. Each type of computer may have its own structure at this level of description. The second level of description will reveal how differently constructed computers have elements, as a result of this causal infrastructure, which can be interpreted in the same way with regard to what is represented, how it is, and what algorithm is embodied in it. Our ability to interpret the activity of computers in

this way is taken to justify one talking of certain parts of the computer as means of representation and of the computer as embodying an algorithm.

Marr's approach to psychological explanation is not idiosyncratic. It is pretty much in accordance with the account offered by Robert Cummins³ and represents something of an orthodoxy. Cummins supposes that a legitimate request for an explanation is 'How is it that system S has capacity C?' This is not taken to be a request for information as to what caused possession of the capacity C, at least, if this is understood to concern some cause, distinct from the capacity, which brought about the possession of the capacity. Instead, the question hopes for information as to what structure the system has such that that system possesses the capacity. For instance, one may ask why does a watch have the capacity to tell the time and the answer which one wishes is not a story about the good watchmaker who laboured to produce the watch but rather how the spring, cogs and wheels fit together so that the time's passing is calibrated in the familiar way.

Given that one allows the legitimacy of the question, Cummins proceeds to offer the following model of explanation. Each capacity that a system has can be broken down into sub-capacities such that if these sub-capacities are combined in a suitable manner within a

system it will follow that the system possesses the capacity. Parts of the system are identified as providing the system with the sub-capacities. We will call these parts 'implementations of a sub-capacity'. The sub-capacity, itself, is broken down into sub-sub-capacities and parts of the implementation of the sub-capacity will be an implementation of a sub-sub-capacity. A full explanation has been provided of how a system possesses a capacity if we can point to all the various parts of the system which implement sub-capacities, sub-sub-capacities and so on, and show that these sub-capacities are combined in a manner suitable for providing the system with the capacity in question.

The talk of capacities and sub-capacities is roughly what Marr would characterise as the computational level of understanding for each component part of the system. In identifying a second level of explanation he supposes that some capacities of an information processing system require an implementation that includes means of representation and the embodiment of an algorithm, and this requirement is taken to be sufficiently of interest and susceptible to independent empirical research that it is also worth discussing separately.

The picture that Marr has offered of what is involved in psychological explanation has been recently complicated by the identification of other levels deemed to be

important.⁴ But we do not need to discuss these complications in the present circumstances. We have said enough in order to appreciate whether it is likely that norm-based explanation will be integrated into psychology.

It is clear that the style of explanation that we have sketched above may be thought by some to be illegitimate, those, in particular, who emphasise the pre-eminence of causal explanation. However, the defence that we offered against such a view in the previous chapter should suffice to justify the present approach to explanation.

2. The positive case for norm-based explanation as a level of psychological explanation.

The positive case for norm-based explanation is as follows. First, the subject matter of cognitive science coincides with that of the particular case of norm-based explanation we discussed in the previous chapter. Both concern the mind, and specifically, its cognitive operations. Second, we will argue that norm-based explanation is similar to that which we have found to characterise part of cognitive science. If this is true, then those who wish to repudiate norm-based explanation while adopting the explanations that we find in cognitive science will be making an unmotivated distinction.

There are some obvious similarities between the type

of norm-based explanation discussed previously and the computational level. Both involve the identification of a goal, and, the identification of various states of a system in terms of their contribution towards obtaining it. In the case of the norm-based explanation, the aim was to believe only true propositions, and the inferential relations between intentional contents indicated one way in which this would be obtained. In the case of the computational level for vision, the goal is to obtain perceptual information about the environment and various processing stages are identified which contribute towards this. There seems, moreover, to be a similarity between the goals selected, broadly speaking, both are concerned with information. Finally, both the computational level and our type of norm-based explanation partly characterise a system in terms of intentional content, and the manipulation of such content. They are not concerned with the means of representation for content.

These similarities seem to suggest that a norm-based explanation that appeals to truth can be seen as potential explanation of cognitive science at about the same level of 'abstraction' from implementation as that which Marr identified as the computational level. However, there may be thought to be at least one salient difference between them. It might be thought that the computational level was concerned to identify capacities that the system

has, whereas, the norm-based explanation identifies components of the system by reference to an aim which the system only to some extent achieves. Loosely, the norm-based explanation talks about that at which the system aims whether or not it has the capacity.

Although it is possible that there is this difference, it is not certain. At least in Marr's own work a measure of idealisation occurs. He takes the visual system to obtain information about the world, without qualification, and then searches for ways in which it might do that. If no idealisation was involved one could not correctly describe the visual system in these terms. One would have to say more precisely what it was capable of obtaining and what it was not. It would be described as a visual system which was only geared to obtain information from the world in certain circumstances, those in which the tacit assumptions that the system makes about the relationship between its stimulation and the distal stimulator are borne out.

Unfortunately, the existence of idealisation in Marr's approach does not, by itself, imply that his theory contains explanations that we may consider to be, or to be analogous to, our norm-based explanation. This matter will depend upon the type of explanatory entities Marr employs, and, about this, a controversy is currently raging between those who believe that he is committed to

holding that the content of perceptions is environment-dependent in some sense, and those who believe he is not.⁵ Obviously, as far as we are concerned, it would be congenial if he were so committed, but now is not the appropriate time to comment on a sophisticated line of argument. It might be that a theory developed in cognitive science pronounces in no clear way upon the character of what we might fairly describe as the notion of a state which has initially been developed in a somewhat more traditional framework. Nevertheless, even if things do not turn out as conveniently as we would wish, we may suggest that the presence of idealisation at the computational level opens up the possibility of a series of norm-based explanations being offered, within the confines of cognitive science, which would make a contribution to our understanding of the system in which we are interested.

Nor should one suppose that psychology is alone in setting this sort of scientific precedent. Evolutionary theory in biology suggests that there is a selection mechanism in nature for those creatures that can best pass forward their genes. But of course, this cannot be strictly correct, because nature does not contain within itself a 'perfect' selection mechanism, it only selects in certain circumstances. There are times when the inadequate will not be weeded out. Nevertheless, when

biologists talk about the function that certain features of organisms have, they only characterise them by their contribution to the reproductive capacity of the organism in circumstances where the selection mechanism is, as one might say, fully operative. So, there is an appeal to a norm for their characterisation similar to that envisaged.

Now, it might be thought that a significant difference is that an ability to reproduce is something one can describe in scientific terms whereas 'truth' is not. Hence, only the former norm is scientifically acceptable. We shall not go into what 'describe in scientific terms' means. Our answer is just that we are a victim of our use of 'truth' as shorthand for the collection of propositions that truly and exhaustively describe the world. The idea, to recall, was that the aim is to believe only true propositions and disbelieve false ones. So long as an account of intentional properties can be provided of the sort we have identified, we can use this account to identify in scientific terms what the subject believes. The world about which human subjects aim to obtain information itself determines the membership of the list of propositions that we used to capture, in figurative terms, the aim of human subjects. So, on the assumption that the world is scientifically describable, we would have a way of specifying both the aim and the relevant components of

the aimer, in scientific terms. We would then be committed to providing a scientific account of the predicate '- is true'.

Even if the account we have put forward of the norm of truth can be made scientifically acceptable, to the extent that we have just suggested is possible, it might be thought that there is another discrepancy between, in this case, the approach adopted by evolutionary theory and our own approach. The problem is that it is relatively easy to specify in scientific terms what reproduction is, and so we have a fair idea of the norm of evolutionary theory, but to specify the norm that we recommend would be to give a true description of the entire universe. One cannot hope to be able to do this.

Although there is this disanalogy, it is not present when one contrasts our approach with the computational level of explanation that Marr puts forward. Marr presumes that he has some understanding of what features the physical world possesses so that he can then specify the computational task of the visual system and its components. But, one might say, a description of all the features of the world that we can perceive is just not available to us. Marr got round this type of problem by using the norm as a specification of what a complete theory of vision would have to provide, to which he was working in a piece by piece way. For instance, he would

focus on how the human visual system may get information about depth. Just such an approach is also available with regard to the type of explanation we have suggested should be offered. The idea would be to use the current state of knowledge about parts of the world to determine what we think is represented by a subject's means of representation and update this in the light of future developments in much the same way as we would update the task we perceived the visual system to compute in the light of developments in our understanding of the world.

Nevertheless, even if one does allow that the norm to which we have appealed is scientifically legitimate and one recognises that some idealisation is present in the types of explanation discussed in this chapter, it still might be thought that our type of explanation is not genuinely scientific. It is far too tolerant in its application. As we have already noted, it may be applied to human subjects even if they hardly behave as if they are orientated towards a particular norm, so long as one of the other conditions are met.

We should recognise that this is a genuine disanalogy. What we shall suggest is that in certain instances the type of explanation that we put forward is scientific in others it is not. If the subject to which it is applied does behave more or less as if he or she is orientated towards a particular norm, then the

explanations provided by norm-based explanation are genuinely scientific. Otherwise, we may view them as 'interpretative'. So, our positive argument for allowing that the norm-based explanation we recommended provides a genuine form of psychological explanation ends with an empirical assertion. Human subjects do behave, to a reasonable extent, as if they are aimed at the norm of truth. It is upon this claim that our defence ultimately rests.

3. The negative case and eliminative materialism

A number of arguments that have been put forward against the ascription of mental events with intentional contents rely upon various claims about what such ascriptions fail to explain. For example, Paul and Patricia Churchland⁶ suggest that sleep, madness and learning cannot be explained merely by citing the relevant subject's beliefs and desires. Equally, they question whether any explanations in terms of these types of entities can be integrated into a more comprehensive explanation including the results of neuroscience. They take these matters to indicate something fundamentally wrong with such ascriptions. However, it is to be hoped that the discussion of the last chapter and this one reveals how we may grant that the ascriptions of beliefs and desires do not fully explain, or explain at all, the

phenomena they mention, yet at the same time provide some sort of explanation that licenses their ascription. Also, it is to be hoped that we have given some indication of how the integration of such ascriptions into a more comprehensive explanation may take place. Thus, given the success of our discussion, the application of the norm-based explanation we outlined is not undermined by the Churchlands' arguments against the ascription of the entities to which it appeals in its explanations.

A second assault on the explanatory status of intentional contents stems from what many see as the plausibility of certain models of cognition, connectionist models.⁷ Unfortunately, only a brief description of such models can be given here. Connectionist models of the mind are computers constructed in such a way that their information processing resembles the pattern of activity at the neural level of the human brain. Inputs are transformed into outputs by travelling through a network of nodes and connections between nodes. The connections are weighted, by 'teaching' the network through trial and error, so that the signal travelling along the connections is boosted or diminished at a node, and, as a consequence, a specific input-output correlation is set up, the one required for a particular information processing task.

Connectionist models are said to be incompatible with allowing intentional contents an explanatory role

because, in allowing an explanatory role to them, we are committed to holding that they are isolable parts of the human subject's mind. This, in turn, is because we take individual intentional contents to figure in the causal explanation of behaviour, and that means we need to distinguish between causally active parts and causally inactive parts of the network at any one time. However, the point about connectionist networks is that in many cases they do not allow that there are isolable parts of the required sort.

Let us suppose, for the sake of argument that a connectionist model is both the correct model of the human mind, and that in it there are no isolable bits to correspond to our ascriptions of individual intentional contents. Would it follow that our ascription of intentional contents in a norm-based explanation is undermined? It is hard to see why. The explanatory role that we have accorded intentional properties is non-causal. So we are not committed to there being isolable elements. On the other hand, we have said that it would only be relevant to cite a certain intentional content in an explanation of behaviour if it was attributed to a cause of the behaviour. Thus, we would be committed to saying that the correctness of the connectionist model would probably imply that more intentional contents were relevant to the explanation of behaviour than we imagined

if certain mental events are genuinely inextricable from each other, when we examine the network. However, there seems nothing unfortunate about this commitment.

The main threat for any explanation based upon intentional content is, according to Stephen Stich⁸, that it lacks certain virtues that scientific explanations should have. First, the conditions of application for these explanations should not be vague. Second, the explanations in question should seek to capture all the interesting generalisations that one can make about the entities to be explained. Third, correct identification of the explanatory entities should not depend upon peculiarities in the context of inquiry, or idiosyncrasies in the identifier's character.

To illustrate how intentional contents fail to satisfy these requirements Stich provides numerous examples. We shall focus on just one type of case, although it is to be hoped our remarks have general application. An individual suffering from chronic worsening memory as he or she gets older, may assert that Kennedy was assassinated, to update his example. The subject utters the words 'Kennedy was assassinated' sometimes just in the course of conversation, sometimes when asked 'What happened to Kennedy?' Nevertheless, in spite of this, the subject is unable to tell us who Kennedy was, whether Kennedy was a man or a dog, whether

assassination was a delousing process or involved death, or even, what death is.

In the absence of these abilities, Stich maintains we would generally find it implausible to assign the relevant intentional content to the subject, expressed by the words cited. But, along the progression from relatively normal memory to the deficits mentioned, it is just not clear when the subject ceases to have the belief, so, the applicability of this explanatory notion is vague.

Moreover, on those occasions when we take the notion of belief not to be applicable, we can no longer explain the utterance of the words 'Kennedy was assassinated' by saying that they were uttered because the subject believed the proposition expressed by them. Instead, we would have to provide an explanation in terms of some other sort of internal state. But, once we recognise the availability of this other explanation, it may appear as if we should always have appealed to it, the explanation in terms of belief being, strictly speaking, redundant. It is in this sense that Stich feels interesting generalisations are lost if we stick with just beliefs and desires, when we can.

Finally, the correct ascription of intentional content is thought to depend both upon context of inquiry and upon the ascriber's perspective. Sometimes we are prepared to say that somebody believes something if they

merely commit themselves to the truth of a certain English sentence. For instance, one might ask the subject, 'What do you believe happened to Kennedy' and take the answer as indicative of his or her beliefs. Alternatively, reflecting, perhaps philosophically, upon whether the subject 'really' believed that Kennedy was assassinated we may decide that he or she does not because he or she fails to possess the concept of assassination. Thus, we see the effects of context. Stich says that there is no further context-free standard of correctness.

The correct ascription of intentional content is thought to depend upon the ascriber because, Stich believes, we specify such contents on the basis of a comparison between ourselves and the person to whom we ascribe the content. For instance, when I state that a subject believes that philosophy is difficult, Stich holds that I am ascribing to a subject a belief state that would play a typical causal role in my utterance of 'Philosophy is difficult', if that utterance had a typical causal history.⁹

The complexity and sophistication of Stich's position renders it impossible to deal adequately with it in the space required. Nevertheless, we can indicate the form that a defence of our own approach would take.

We are committed to there being a fact of the matter as to whether or not a subject possesses an intentional

property and we have said that one may determine the character of the fact by investigation of the subject's constitution on those occasions when we are most confident in our ascription of the intentional property. Nothing that Stich says rules out there existing such circumstances. Given that we can determine the relevant facts, while the ascription of content may be context sensitive, the correctness of the content would not be. Of course, there are likely to be a number of vague cases in which it is not clear whether we should count a particular set of facts as constitutive of the occurrence of intentional properties or not. But, this situation has been faced with all the sciences with regard to the entities that are classified by them. So, we should be untroubled by this. Our investigations into the workings of a subject's mind are likely to suggest a useful resolution of the application of our taxonomy of intentional contents, a matter which, of course, we should not prejudge.

We may concede to Stich that subjects do, on the whole, ascribe to others intentional contents on the basis of a comparison with themselves. Nevertheless, it does not follow from this that when a subject is correct in his or her ascription there is no fact which constitutes this correctness apart from a similarity between the ascriber and the person to whom the intentional contents were

ascribed. We are committed to saying that there is something else that makes the ascription correct. Thus, the correct ascription of intentional contents is not dependent upon the ascriber's perspective.

Stich's final objection was that interesting generalisations in cognitive science would be lost if we stuck only with the ascription of intentional contents and, indeed, that certain creatures that could otherwise be fruitfully compared with us could not be the subject of its investigation at all. We can agree with him that this is so, but we are not recommending that we do stick with just this strategy. Instead, it is taken to provide a high level of explanation under which there are many others. So it seems that we can accuse Stich of missing out on generalisations and not us. Our admission that there are levels of explanation below that which we have categorised as norm-based allows us to capture all the generalisations he would wish us to, and we have an additional number on top. Equally, the failure of certain norm-based explanations to apply to some creatures is not a criticism. There are certain preconditions for their successful use. This, of course, rules their application out in some situations, but also means that when they do apply they capture facts of the situation which might otherwise have been missed.

We have now come to an end of our assessment of whether the norm-based explanation we described may be integrated into psychology. We suggested that there were some grounds for thinking so, and discerned no argument that was sufficiently devastating to undermine our optimism. Of course, it is hard to be certain in these sorts of cases as no one paradigm of psychological explanation is both pre-eminent and fully articulated. Nevertheless, the one we chose to discuss has been pretty influential and it seemed of interest to draw attention to certain similarities between it and the norm-based explanation we identified.

References

1. David Marr (1982), Vision (Ch.1)
2. The example is Marr's.
3. R. Cummins (1983), The Nature of Psychological Explanation, (1975), 'Functional Analysis', Journal of Philosophy 72.
4. Z. Pylyshyn (1984), Computation and Cognition;
C. Peacocke (1986), 'Explanation in Computational Psychology: Language, Perception, and Level 1.5', Mind and Language.
5. T. Burge (1986), 'Individualism and Psychology' The Philosophical Review 95, (1986), 'Cartesian Error and the Objectivity of Perception' (P. Pettit and J. McDowell eds., Subject, Thought and Context); G. Segal (1989), 'Seeing what isn't there', The Philosophical Review 98; M. Davies (1991) 'Individualism and Perceptual Content', Mind 100.

6. P.S. Churchland (1980), 'A Perspective on Mind - Brain Research' Journal of Philosophy 77; P.M. Churchland (1981), 'Eliminative Materialism and the Propositional Attitudes' Journal of Philosophy 78.
7. W. Ramsey, S. Stich and J. Garon (1990), 'Connectionism, Eliminativism and the Future of Folk Psychology' (J. Tomberlin, ed., Philosophical Perspectives 4). For discussion of this article, see A. Clark (1990), 'Connectionist Minds' Proceedings of the Aristotelian Society 90.
8. S. Stich (1983), From Folk Psychology to Cognitive Science. We have no time to consider the other arguments that Stich provides for the conclusion that beliefs, desires and so on do not exist. However, in general, we should agree with the discussion and rejection of these arguments in T. Hogan and J. Woodward (1985), 'Folk Psychology is Here to Stay' The Philosophical Review 94.
9. S. Stich (1983) p.88.

Conclusion

Our aim was to try to assess the likely role that mental events will play in psychological explanation. We suggested that there were two arguments that were particularly threatening: one based upon the causal closure of the physical world; and, one based upon the conflict between the environment-dependence of some mental events and the principle that psychology should cite only entities which were not so dependent, in the explanation of behaviour. It is time to draw together our assessment of these arguments, beginning with the second.

We suggested that the intuition on which the second argument rested was the following:

Psychological explanation is just causal explanation. We have shown that if this were true, the environment-dependent aspects of any mental events would have no role to play in psychological explanation because these aspects are causally epiphenomenal. On the assumption that most, if not all, the intentional content we ascribe is environment-dependent, it would follow that the explanation of a piece of behaviour we give by citing mental events with intentional content does not depend for its explanatory worth on the intentional contents of the

mental events. Thus, whatever we may think to the contrary, it is not true that the intentional contents are explanatory in the following explanation: I opened my door because I believed that there was a noise outside and wanted to find out what it was.

Now, it is possible just to deny that the mental events such as beliefs and desires have environment-dependent intentional content. However, the approach we chose to adopt assumed that this option was not open to us so that we may assess whether there were other ways in which we could resist the epiphenomenalist challenge. Instead, we argued that although we should accept that psychological explanation was a species of causal explanation, it has other aspects to it as well. It was to these aspects that we appealed in order to provide specific explanatory roles for mental events categorised by their environment-dependent intentional content. In the previous chapter, we tried to defend the explanatory roles we had identified against the claim that they were not part of an explanation that should be offered by scientific psychology. We did this by showing that the style of explanation that one finds in cognitive psychology is similar to that style of explanation, norm-based explanation, in which the environment-dependent mental events had an explanatory role. Although there are reservations one may have about the explanatory status of

norm-based explanation, those we discussed in Chapter 9, many of the same reservations would apply to the types of psychological explanation cited in Chapter 10. So, there is no special reason to deny that environment-dependent mental events have a role to play in psychological explanation.

Naturally, if it turned out that, contrary to what we supposed, intentional properties are phenomenal properties, the final conclusion to which we come, regarding whether or not they are explanatorily epiphenomenal, depends upon what we say with regard to the first argument we considered. With this in mind, we should endorse the following conditional.

If intentional properties are not phenomenal properties, then intentional properties are not explanatorily epiphenomenal.

The situation in which we have found ourselves becomes considerably less happy when we review the conclusions that we came to with regard to the argument concerning the causal closure of the physical realm. We argued that the phenomenal properties of mental events were non-physical. If the physical realm is causally closed, these properties will be causally epiphenomenal with regard to human behaviour. We have, of course, not considered whether they may be causes of mental events which themselves do not give rise to behaviour because it

was thought that the consequences of the argument for the causation of behaviour were sufficiently counter-intuitive to render the argument worthy of scrutiny.

The phenomenal properties constitute, by definition, the appearance of a subject's mental life. If the argument is sound, it follows that whenever we feel that we are acting as a causal consequence of features of our mental life that we apprehend we are under an illusion, for these features are causally epiphenomenal. This is surely a rebarbative conclusion.

One way in which we might avoid it is if we challenged our account of the nature of the physical. For instance, we might argue that it is not obvious that something is not physical if its existence is metaphysically sufficient for an awareness of it. However, one may then wonder what is the issue which divides those who believe that mental properties are not physical from those who do not. Any feature of the mental that is pointed to as distinctively non-physical is embraced in an expanded conception of the physical. It begins to look as if those who believe that everything is physical are right, but only because they believe that the physical is everything. The issue, if it exists at all, would become a disagreement between those who believe that the mental will always be mysterious, and those who think there will be a scientific account of it.

Another option would be to reject the claim that there are physical things and adopt some version of idealism. Obviously, then the dichotomy upon which the argument rests would vanish. Nothing that we have said rules out such an approach, however, it is worth considering whether we may provide a solution to our problem given that physical things do exist roughly in the way we envisage. Our success or otherwise would obviously have ramifications for the appeal of idealism.

There seem to be two alternatives. First, we could deny that the physical realm is causally closed and, further, claim that we will be able to find traces of the causal activity of phenomenal properties when we obtain a better understanding of the brain. In support of this position, it might be held that some of the arguments in favour of the causal closure of the physical realm are less compelling when we consider how they would apply in the case of the mental. For instance, one of the constraints that was put forward to rule out the explanation of phenomena by non-physical entities was that we should not have to alter the type of explanation we offer depending upon whether or not the phenomena were observed. However, in the case of the explanation of behaviour by the phenomenal properties of mental events, there is always going to be an observer, the subject to whose mental life we refer. So the problem does not

arise. The explanatory entities we put forward would be applicable in all relevant cases.

The second argument we put forward in favour of the principle that the physical realm is causally closed rested upon the claim that it is massively unlikely that phenomenal properties need to be referred to, to explain the occurrence of behaviour, since there is no sign, in the workings of the brain, that there is a gap in the causal story that they are needed to fill. However, given the fact that we have very good, albeit introspective, evidence that there is a causal relationship between phenomenal properties and behaviour this claim begins to look less convincing. We may suppose that a causal relationship does exist between the non-physical and the physical and that current neurophysiological research has simply not adopted the right technique to identify its occurrence. However, these reflections are apt to seem inconclusive.

A more general point that we should make against our previous endorsement of the 'closure' principle is that any scientific justification of it is significantly undermined once we allow that phenomenal properties are non-physical. The principle then clashes with another. It is this.

There must be a reason why any occurrence of a property occurs at the spatio-temporal location it

does.

We may call this a principle of sufficient reason for properties. It is standard practice in science to suppose that the 'reason' in question is a causal one. Applying this principle to the case of our non-physical phenomenal properties we must allow that something caused their occurrence. Now the cause in the last resort cannot just be other non-physical things, for we want to know why, in particular, collections of phenomenal properties are tied to particular brains. Either we must suppose that there is a whole system of mysterious non-physical properties, for which we have never had evidence up until now, to make the phenomenal properties occur in the appropriate places, or we must allow that physical properties do cause non-physical properties. Since the first possibility makes the whole metaphysical picture to which we seem committed even more unattractive, it looks as if we may be forced to accept the latter possibility.

One might try to limit its damage by the following reformulation of the principle of the causal closure of the physical realm. It now reads.

(1)''The physical realm is causally closed, in other words, no physical entity will be caused by some other entity as a consequence of the latter's possession of a non-physical property.

The principle leaves open the possibility that the

physical may cause the non-physical but rules out the reverse. If it were true, we would have an explanation of why we do not see evidence of the workings of non-physical properties.

If the principle just put forward was acceptable then it might be thought to follow that phenomenal properties are causally epiphenomenal in spite of the fact that there is a causal relationship between them and physical properties. To obtain this result, we would of course have to reformulate our account of causal epiphenomenalism so that something counts as causally epiphenomenal if and only if it is not a cause. Some may view the reformulation as long over due. However, the legitimacy of the principle depends upon whether we can justify a belief in the objective existence of causal priority. If one allows that we have reason to believe that there is an asymmetry between cause and effect of the sort to which our understanding of the notion of causal priority appeals, then it may be that it is significant to distinguish between the principle with which we began our inquiry, and our current formulation. The necessity of causality, so to speak, flows from physical to non-physical but not vice versa so as to disturb the workings of the physical. However, we saw that it is notoriously hard to justify, or even understand, the notion of causal priority which is required. Such asymmetry as there is

between cause and effect often seems something we put into the world by our approach to it, and not something that is in the world.

If it turns out that there is no objective necessity corresponding to our notion of causal priority, then the reformulation of the 'closure' principle will be futile. Calling something a cause and something else an effect would not detract from the basic fact that the non-physical has just as much influence upon the physical as the physical has upon the non-physical.

Even if we do allow that there is an objective difference between the cause and the effect, it is not clear how we can be sure that the causal priority works in our favour. Perhaps the occurrence of non-physical mental properties are causally prior to the occurrence of the physical properties with which they are correlated. Initially, this might seem an implausible suggestion. It may be thought that if, as seems to be the case, brains were composed from the same type of physical properties as occur elsewhere, then to admit that non-physical properties were causes of these properties would be to admit that what might count as sufficient to bring about the occurrence of a physical property outside the brain, would not be inside. The occurrence of a non-physical property would also be required. Thus, we would be able to see evidence once more of the workings of the non-

physical, and, for all that we have said, this is still unlikely. Unfortunately, matters are not quite as simple as this line of reasoning would suggest. To see this we should consider what follows if we see no evidence of the activity of non-physical properties.

Suppose, contrary to what we have held up until now, we find that, however hard we try, we cannot find evidence of the causal efficacy of phenomenal properties except of the sort we have already indicated that we obtain through introspection. What should we then conclude? There are two possibilities. First, we may hold that our non-introspective access to the workings of our brain is just not the sort of access that enables us to isolate the relevant causal relationship and recognise both relata. This would not just be a claim about our access up until now, but would be a claim about the very nature of non-introspective access. Surely this would be more rational than to conclude that our properties are causally epiphenomenal?

However, it is the other way in which the failure of the principle would not be detectable which is of relevance to the point just made earlier. If one allows that phenomenal properties are causally correlated with physical properties, and that the physical behaviour of the brain is what one would expect and predict from its physical constitution, then one must suppose that the

causal correlation between non-physical and physical renders the causal relations between physical properties, that are usually sufficient, insufficient. The occurrence of non-physical properties invisibly alters the workings of the causal relation between occurrences of physical properties. One physical property is only sufficient for another physical property in the brain, in conjunction with a non-physical property which is directly causally related to the first physical property. Yet, outside the brain the first physical property is sufficient for the second.

This is no doubt a puzzling view but it would be an alternative way of accommodating most of our intuitions. Phenomenal properties would be causally efficacious, and the physical realm would be predictively closed. Scientists need not identify anything but occurrences of physical properties in their attempts to predict future physical events. It is just that predictive closure does not imply causal closure.

The alternative to abandoning the 'closure' principle is that we account for the intuition that phenomenal properties are among the causal antecedents of behaviour by supposing that phenomenal properties are metaphysically, but not causally, connected to physical properties. We would have obtained, by reasoning alone, an insight into a mysterious connection between the

phenomenal and physical which, so to speak, holds our intuitions together. The metaphysical connection between phenomenal properties and physical properties is such that the former could not occur without an appropriate occurrence of the latter, which results in the illusion that the phenomenal properties have causal powers, when in fact they do not. We would also have, by this approach, an alternative way in which we could satisfy the principle of sufficient reason for properties. It is perhaps worth pointing out that the relationship in question could not be one of mere supervenience as this would be incompatible with the 'closure' principle. If the argument we offered in Chapter 3 is correct, then such supervening properties are causally efficacious. Thus, we would have to allow that the occurrence of phenomenal properties caused physical events after all, contrary to the 'closure' principle. Since the last alternative is not particularly satisfactory except for those who love philosophical mysteries, the weight of evidence really does favour rejecting the 'closure' principle.

If it turned out that there were better arguments than we have uncovered for the 'closure' principle and there is an objective causal asymmetry, we might have to face the possibility that either phenomenal properties are causally epiphenomenal, or we have a radical misunderstanding of what awareness is. One way in which

we could soften the blow is by providing a non-causal explanatory role for these properties. An obvious suggestion would be that they are a familiar means of identifying what really does the causal work.

Unfortunately such a response is likely to seem unsatisfactory. It might be said that our intuition concerning the explanatory role of phenomenal properties is that they are more explanatorily significant than allowed by our surrogate role. The same dissatisfaction may well plague our conciliatory approach to the explanatory role of environment-dependent intentional content. This would seem a fair criticism. We should not be fobbed off with fake explanatory roles and told that this indicates that certain challenges of epiphenomenalism are not really challenges at all.

Nevertheless, we have attempted to provide some sort of reply to the challenges. We have in one case attacked the emphasis placed upon causal explanation at the expense of all others. It may be possible that this approach is available elsewhere. Any epiphenomenalist challenge must ultimately rest upon intuitions about the significance of an entity for the explanation of other entities. If we undermine a certain way of understanding this significance, then the possibility opens up of recognising a certain amount of causal epiphenomenalism, but learning to live with it.

We focused on phenomenal properties and intentional properties in the assessment of our question, concerning the likely role of mental events in psychological explanation, because it seemed that the greatest threat to their role came from claims about the explanatory worth of these properties. We have tried to show that the threats mentioned are not so great. Thus, we may say that many of the reasons for doubting that mental events will play such a role can be resisted. Our intuition that mental events are likely to play a role in psychological explanation still stands, and this, it is reasonable to suppose, is the conclusion to which we wished to come.

Bibliography

Aristotle

- De Amina (in J. Barnes (ed.), The Complete Works of Aristotle, Princeton University Press)

D.M. Armstrong

- (1968) A Materialist Theory of Mind (Routledge & Kegan Paul)
- (1980) The Nature of Mind (Cornell University Press)
- (1988) What is a Law of Nature? (Cambridge University Press)

D.M. Armstrong and N. Malcolm

- (1984) Consciousness and Causality (Basil Blackwell)

A.J. Ayer

- (1973) The Central Questions of Philosophy (Penguin)

J. Bacon

- (1986) 'Supervenience, Necessary Coextension and Reducibility', Philosophical Studies 44

J. Bennett

- (1988) Events and their Names (Oxford University Press)

S. Blackburn

- (1984) 'The Individual Strikes Back', Synthese 58
- (1985) 'Supervenience Revisited (I. Hacking, ed. Exercises in Analysis, Cambridge University Press)

N. Block

- (1980) Readings in the Philosophy of Psychology Vol.1 (Methuen)
- (1980) 'Are Absent Qualia Impossible?', Philosophical Review 89
- (1987) 'Can the Mind change the World?' (in G. Boolos (1990), ed., Meaning and Method, Cambridge University Press)

F. Brentano

- (1874) Psychology from an Empirical Standpoint (Routledge & Kegan Paul)

T. Burge

- (1979) 'Individualism and the Mental' (P. French, T. Uehling and H. Wettstein, eds., Midwest Studies in Philosophy Vol. IV, University of Minnesota Press)
- (1982) 'Other Bodies' (A. Woodfield, ed., Thought and Object, Oxford University Press)
- (1982) 'Two Thought Experiments Reviewed', Notre Dame Journal of Formal Logic Vol.23

- (1986) 'Individualism and Psychology', Philosophical Review 95
- (1986) 'Intellectual Norms and Foundations of Mind', Journal of Philosophy 83
- (1986) 'Cartesian Error and the Objectivity of Perception (P. Pettit and J. McDowell, eds., Subject, Thought and Context, Oxford University Press)
- (1989) 'Individuation and Causation in Psychology', Pacific Philosophical Quarterly 70

- R. Carnap,
 - (1928) The Logical Structure of the World (University of California Press)

- R. Chisholm
 - (1957) Perceiving (Cornell University Press)

- N. Chomsky
 - (1969) "Quine's Empirical Assumptions" (D. Davidson and J. Hintikka, eds., Words and Objections, D. Reidel)

- P.M. Churchland
 - (1979) Scientific Realism and the Plasticity of Mind (Cambridge University Press)
 - (1989) A Neurocomputational Perspective (The MIT Press)

- P.S. Churchland
 - (1980) 'A Perspective on Mind-Brain Research', Journal of Philosophy 77
 - (1986) Neurophilosophy (The MIT Press)

- A. Clark
 - (1989) Microcognition (The MIT Press)
 - (1990) 'Connectionist Minds', Proceedings of the Aristotelian Society 90

- G. Cohen
 - (1978) Karl Marx's Theory of History (Oxford University Press)

- E. Craig
 - (1982) Meaning, Use and Privacy', Mind 91
 - (1987) The Mind of God and the Works of Man (Oxford University Press)

- R. Cummins
 - (1975) 'Functional Analysis', Journal of Philosophy 72
 - (1983) The Nature of Psychological Explanation (The MIT Press)

- D. Davidson
 - (1980) Essays on Actions and Events (Oxford University Press)

- (1984) *Inquiries into Truth and Interpretation* (Oxford University Press)
 - (1985) 'Reply to Harry Lewis' (in B. Vermazen and M.B. Hintikka, eds., *Essays on Davidson, Actions and Events*, Oxford University Press)
- M. Davies
- (1986) 'Individualism and Supervenience', *Proceedings of the Aristotelian Society, Supplementary Volume 60*
 - (1991) 'Individualism and Perceptual Content', *Mind* 100
- D.C. Dennett
- (1979) *Brainstorms* (Harvester Press)
 - (1987) *Intentional Stance* (The MIT Press)
- R. Descartes
- (1644) *The Principles of Philosophy*
 - (1641) *Meditations on First Philosophy* (both in *The Philosophical Writings of Descartes* (2 Vols.), translated by J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge University Press)
- F. Dretske
- (1981) *Knowledge and the Flow of Information* (Basil Blackwell)
 - (1986) 'Misrepresentation' (R. Bogdan, ed., *Belief*, Oxford University Press)
 - (1988) 'The Explanatory Role of Content' (D. Merrill and R. Grimm, eds., *Contents of Thought*, University of Arizona Press)
 - (1988) *Explaining Behaviour* (The MIT Press)
 - (1989) 'Reasons and Causes' (J. Tomberlin, ed., *Philosophical Perspectives Vol.3*, Ridgeview Publishing Company)
- C.J. Ducasse
- (1951) *Nature, Mind and Death* (Open Court Publishing Company)
- M.A.E. Dummett
- (1976) 'What is a Theory of Meaning II' (in G. Evans and J. McDowell, eds. *Truth and Meaning*, Oxford University Press)
 - (1978) *Truth and Other Enigmas* (Duckworth)
 - (1982) 'Realism', *Synthese* 52
- G. Evans
- (1982) *The Varieties of Reference* (Oxford University Press)
- P.K. Feyerabend
- (1963) 'Materialism and the Mind-Body Problem', *The Review of Metaphysics* Vol.17

- (1963) 'Mental Events and the Brain', Journal of Philosophy 60
- J. Fodor
- (1975) The Language of Thought (Harvester Press)
 - (1980) 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology', The Behavioural and Brain Sciences Vol.3
 - (1987) Psychosemantics (The MIT Press)
 - (1990) A Theory of Content (The MIT Press)
 - (1991) 'A Modal Argument for Narrow Content', Journal of Philosophy 88
- B. van Fraassen
- (1970) An Introduction to the Philosophy of Time and Space (Columbia University Press)
- G. Frege
- Collected Papers (edited by B. McGuinness) (Basil Blackwell)
- N. Goodman
- (1976) Languages of Art (Hackett Publishing Company, Inc.)
- W.D. Hart
- (1988) The Engines of the Soul (Cambridge University Press)
- C.G. Hempel
- (1965) Aspects of Scientific Explanation (The Free Press)
- T. Hogan and J. Woodward
- (1985) 'Folk Psychology is here to stay', The Philosophical Review 94
- T. Honderich
- (1982) 'Against Teleological Historical Materialism', Inquiry 25
 - (1988) A Theory of Determinism (Oxford University Press)
- J. Hornsby
- (1986) 'Physicalist Thinking and Conceptions of Behaviour' (P. Pettit and J. McDowell, eds., Subject, Thought and Context, Oxford University Press)
- D. Hume
- (1739/40) A Treatise of Human Nature
 - (1748) An Enquiry Concerning Human Understanding (both edited by L.A. Selby-Bigge, Oxford University Press)

Frank Jackson

- (1977) Perception (Cambridge University Press)
- (1982) 'Epiphenomenal Qualia', Philosophical Quarterly 32
- (1986) 'What Mary Didn't Know', Journal of Philosophy 83

F. Jackson and P. Pettit

- (1988) 'Functionalism and Broad Content', Mind 97
- (1990) 'Program Explanation, A General Perspective', Analysis 50

J. Kim

- (1971) 'Materialism and the Criteria of the Mental', Synthese 22
- (1978) 'Supervenience and Nomological Incommensurables', American Philosophical Quarterly 15
- (1979) 'Causality, Identity and Supervenience in the Mind-Body Problem', Midwest Studies in Philosophy IV (University of Minnesota Press)
- (1984) 'Epiphenomenal and Supervenient Causation', Midwest Studies in Philosophy IX (University of Minnesota Press)
- (1984) 'Concepts of Supervenience', Philosophy and Phenomenological Research 45
- (1987) '"Strong" and Global Supervenience Revisited', Philosophy and Phenomenological Research 48
- (1989) 'The Myth of Non-reductive Materialism', Proceedings of the American Philosophical Association Vol.63

S. Kripke

- (1972/1980) Naming and Necessity (Basil Blackwell)
- (1984) Wittgenstein, On Rules and Private Language (Basil Blackwell)

A. Levison

- (1983) 'An Epistemic Criterion of the Mental' Canadian Journal of Philosophy 13

E. LePore and B. Loewer

- (1987) 'Mind Matters', Journal of Philosophy 84

D. Lewis

- (1972) 'Psychophysical and Theoretical Identification', Australian Journal of Philosophy 50
- (1983) Philosophical Papers Vol.1 (Oxford University Press)
- (1986) Philosophical Papers Vol.2 (Oxford University Press)
- (1983) 'New Work for a Theory of Universals' Australian Journal of Philosophy 61

- B. Loar
- (1987) 'Subjective Intentionality',
Philosophical Topics 15.
- J.L. Mackie
- (1974)/80) The Cement of the Universe (Oxford
University Press)
- D. Marr
- (1982) Vision (W.H. Freeman and Company)
- J. McDowell
- (1978) 'Physicalism and Primitive Denotation, Field on
Tarski', Erkenntnis 13
 - (1981) 'Anti-Realism and the Epistemology of
Understanding' (H. Paret and J. Bouveresse, eds.,
Meaning and Understanding, De Gruyter)
- C. McGinn
- (1989) Mental Content (Basil Blackwell)
- H. Mellor
- (1976) 'In Defense of Dispositions', Philosophical
Review 83
- R. Millikan
- (1984) Language, Thought and other Biological
Categories (The MIT Press)
- T. Nagel
- (1979) Mortal Questions (Cambridge University Press)
 - (1986) The View from Nowhere (Oxford University Press)
- D. Papineau
- (1989) Reality and Representation (Basil Blackwell)
- C. Peacocke
- (1981) 'Demonstrative Thought and Psychological
Explanation', Synthese 49
 - (1983) Sense and Content (Oxford University Press)
 - (1986) 'Explanation in Computational Psychology:
Language, Perception and Level 1.5', Mind and Language
Vol.1
- U.T. Place
- (1956) 'Is Consciousness a Brain Process?', British
Journal of Psychology 47
- H. Putnam
- (1975) Mind, Language and Reality (Cambridge University
Press)
 - (1981) Reason, Truth and History (Cambridge University
Press)

- Z. Pylyshyn
 - (1984) Computation and Cognition (The MIT Press)
- W.V. Quine
 - (1960) Word and Object (The MIT Press)
 - (1974) The Roots of Reference (Open Court Publishing Company)
- A. Quinton
 - (1973) The Nature of Things (Routledge and Kegan Paul)
- W. Ramsey, S. Stich and J. Garon
 - (1990) 'Connectionism, Eliminativism and the Future of Folk Psychology' (J. Tomberlin, ed., Philosophical Perspectives 4, Ridgeview Publishing Company)
- H. Robinson
 - (1982) Matter and Sense (Cambridge University Press)
- R. Rorty
 - (1965) 'Mind-Body Identity, Privacy and Categories', The Review of Metaphysics 19
 - (1965) 'In Defence of Eliminative Materialism', The Review of Metaphysics 19
 - (1980) Philosophy and the Mirror of Nature (Basil Blackwell)
- B. Russell
 - (1912-13) 'On the Notion of Cause', Proceedings of the Aristotelian Society 13
- G. Ryle
 - (1949) The Concept of Mind (Penguin)
- J.R. Searle
 - (1983) Intentionality (Cambridge University Press)
- G. Segal
 - (1989) 'The Return of the Individual', Mind 98
 - (1989) 'Seeing what isn't there', The Philosophical Review 98
- W.S. Sellars
 - (1963) Science, Perception and Reality (Routledge and Kegan Paul)
 - (1966) 'Reply to Aune' (H. Casteneda, ed., Intentionality Minds and Perception, Wayne State University Press)
 - (1968) Science and Metaphysics (Routledge and Kegan Paul)
 - (1969) 'Language as Thought and as Communication', Philosophy and Phenomenological Research 29

- (1975) 'The Adverbial Theory of the Objects of Sensation', *Metaphilosophy* 6

- S. Schiffer
 - (1978) 'The Basis of Reference', *Erkenntnis* 13

- S. Shoemaker
 - (1984) *Identity, Cause and Mind* (Cambridge University Press)

- J.J.C. Smart
 - (1959) 'Sensations and Brain Processes', *Philosophical Review* 68
 - (1963) *Philosophy and Scientific Realism* (Routledge and Kegan Paul)
 - (1963) 'Materialism' *Journal of Philosophy* 60

- S. Stich
 - (1978) 'Autonomous Psychology and the Belief-Desire Thesis', *The Monist* 61
 - (1983) *From Folk Psychology to Cognitive Science* (The MIT Press)

- G. Strawson
 - (1989) *A Secret Connexion* (Oxford University Press)

- R. Taylor
 - (1966) *Action and Purpose* (Prentice Hall Inc.)

- M. Tye
 - (1989) *The Metaphysics of Mind* (Cambridge University Press)

- L. Wittgenstein
 - (1953, 1958) *Philosophical Investigations* (Basil Blackwell)