Chapter 1 Supervenience and Identity

1.1 Introduction

Like many other contemporary philosophers, I have strong physicalist intuitions. I am inclined to think that chemical phenomena, for example, are all at bottom physical, even though chemists do not describe those phenomena in physical terms. What is more, I am inclined to think the same about the phenomena studied by meteorology, biology, psychology, sociology and the other so-called "special sciences".

My aim in this initial chapter is to see how far such physicalist intuitions can be supported by serious arguments. This question is not as much discussed in the contemporary philosophical literature as it might be. Of course many philosophers with physicalist inclinations have formulated different possible versions of physicalism, and explored the relations between them. And many other philosophers, with opposed inclinations, have elaborated various non-physicalist views of psychology, biology, sociology, and other special phenomena. But for the most part neither party has paused to argue its case against the other. The friends of physicalism tend simply to start with their physicalist intuitions, and try to develop a theory which fits them. Their opponents dismiss those intuitions out of hand as symptoms of an overblown admiration for science.

Not all philosophers treat physicalism as beyond debate in this way. An increasing number of contemporary thinkers are coming to recognize that there are plenty of pertinent arguments that bear on the issue.1 Dogmatic physicalists and anti-physicalists alike will do well to attend to these arguments. Anti-physicalists will discover that physicalism is supported by premises which are difficult to deny, even if you have little regard for science. And physicalists will find out why some versions of physicalism are defensible, while others are not.

1.2 Supervenience

Let me start by trying to be a bit more precise about what I mean by physicalism. One simple way of formulating physicalism would be to require that all special properties, like chemical, or biological, of psychological properties, should be identified as types with physical properities, in the way that the property of being hydrogen, say, can be identified with the physical property of having atoms with one proton and one electron. But while such "type identities" may be available within basic chemistry, they seem unlikely to characterize the other special sciences. In particular, it seems unlikely that psychological properties, such as being worried about the future, for example, can be identified with any specific physical properties, along the lines of having a certain arrangement of molecules in your head. It is surely implausible to suppose that all the different people who have ever been worried about the future must have some intra-cranial molecular property in common. And, if that is not implausible enough, what about the future brain-injured people who will have their damaged parts replaced by miracles of silicon-based micro-technology, or the hominid but silicon-based denizens of Proxima Centauri's third planet? Presumably they will be able to worry about the future too. But they can't possibly share

molecular arrangements with the rest of us, given that we don't have any silicon in our brains.

Fortunately for physicalism, type identity is not the only way in which special properties can be viewed as essentially physical. An alternative way of formulating physicalism is in terms of the supervenience of the special on the physical. Supervenience on the physical means that two systems cannot differ chemically, or biologically, or psychologically, or whatever, without differing physically; or, to put it the other way round, if two systems are physically identical, then they must also be chemically identical, biologically identical, psychologically identical, and so on.

The advantage of formulating physicalism in terms of supervenience is that, unlike type identity, this doesn't require that the same physical property must determine a given special property whenever it is instanced. My worrying about the future might involve one molecular arrangement, an arrangement such that that anybody who has it will be worrying about the future; your worrying about the future might be ensured by a different physical arrangement, but again one that suffices to determine that all its possessors are worrying about the future; future brain-damaged patients and Proxima Centaurians will have yet different such physical arrangements; and so on.

How satisfactory an explication of physicalism is the requirement of supervenience on the physical? I shall consider first whether supervenience is necessary for physicalism, second whether it is sufficient.

On the face of it, supervenience seems an obvious necessary condition for physicalism in any given area: if two chemical systems, say, can differ, even though they are physically identical, then it would seem to follow that they must contain something non-physical.

However, an immediate qualification is needed. Suppose two chemical samples are physically identical: they contain exactly the same molecules and have exactly the same internal structure. Nevertheless one may be heavier than the other, if one is one the earth and the other on the moon. So the heaviness of chemical systems does not supervene on their physical characteristics. Yet presumably we don't want on this account to regard physicalism as refuted by the heaviness of chemical samples. If anything supervenes on physical characteristics, surely heaviness does.

The obvious response to this problem is to note that heaviness is a relational property of chemical samples, depending not only on the intrinsic features of the sample, but also on the features of another system, namely, the surrounding gravitational field. Accordingly, we should modify the requirement of supervenience, for relational properties, so as to demand that such properties should supervene, not on the internal physical characteristics of the system at issue, but rather on those plus the physical characteristics of the relevant related system. If we do this, then the heaviness of chemical samples is no longer a counter-example to physicalism: for the heaviness of a chemical sample obviously does supervene on the internal physics of the sample plus the physics of the surrounding gravitational field. (Equivalently, if less naturally, we could say that the relational properties of a system were not really properties of that system as such, but only of some larger system incorporating the

relevant related system, and then require that such relational properties supervene on the physical properties of the larger system.)2

Given this qualification about relational properties, I shall take it henceforth that supervenience is a necessary condition for physicalism. But is supervenience sufficient for physicalism? This is a rather more tricky issue. In outline, we can see how supervenience might suffice. Supervenience says that, if two systems are physically identical, then they must also be chemically identical, biologically identical, psychologically identical, and so on. That is, the shared physical features of these systems determine their special features. But how could this be so, if anything non-physical were required for those special features?

Some care is needed, however, to make this line of thought watertight. The issue depends on exactly how we understand supervenience, and in particular on how strongly we read the "determine" in "the shared physical features of these systems determine their special features". In due course we shall see that there is a weak reading of this "determine" on which supervenience clearly does not suffice for physicalism, and a stronger reading on which supervenience does provide a satisfactory characterization of physicalism. But let me not pause for these technicalities at this point. My primary interest in this chapter, as I said, is not with the characterization of physicalism as such, but with the possibility of arguments which support physicalist views. In line with this, it will make more sense for me to fill in the details of what I mean by physicalism once we have see what arguments are available, rather than before.

Perhaps it will be helpful to be graphic for a moment. The world recognized by physicalism is at bottom a world consisting of physical facts, of particles and fields in motion through space. At this basic level all facts can be described by strictly physical terminology, like "mass", "energy", and "position". However, physicalism, as I am thinking of it, will also allow that we often use non-physical terminology, like "sulphuric acid", "thunderstorm", "elephant", and "thinking of the future", to group and categorize large-scale arrangements of physical facts. Moreover, physicalism allows that such special terminology isn't just a shorthand for complex physical properties: for, in those cases where type identity fails, special categories cannot even in principle be specified in physical terms. Nevertheless, physicalists will say, the instances of any such special kind will still just be complexes of physical stuff. For supervenience, in an appropriately strong sense, implies that nothing more is required for any special kind to be instanced than the physical facts should be thus-and-so. After all, if anything more were required, then presumably it would be possible for the special features of two systems to differ even though they were physically identical, which is just what supervenience rules out.

So far we have been concerned only with what physicalism says. It remains to consider whether we should believe it. In the rest of this chapter I shall argue that physicalism is strongly supported by an important feature of physical science, namely, the internal completeness of physics. However, before proceding, it will be helpful briefly to consider a number of further preliminary points that may be worrying some readers.

1.3 More Preliminaries

1.3.1 Some of you may feel uneasy about my brisk dismissal of the possibility of type identities between physical categories and special categories. In particular, you may feel that if a special subject matter is scientific enough to contain projectible laws, then it would be surprising if its categories were not type identifiable with physical categories. For why should we expect special categories to conform to any stable regularities, if they are determined by different physical structures on different occasions?

I think this question points to a powerful, though not inescapable, argument for type identity, and shall devote the next chapter to it. In this chapter, however, I shall focus on the prior issue of whether we should accept physicalism, understood in terms of supervenience. Once we have decided this, we can then turn to the further issue of whether we should accept type identity as well. My eventual conclusion will be that type identity holds for some, but not all special sciences: more specifically it holds for those special sciences that lack a teleological underpinning.

1.3.2 In this chapter, and in much of the rest of this book, I shall speak as if our "common sense psychology", which attributes beliefs, desires and other familiar states to people, is a "special science". But this is of course a contentious assumption. Many philosophers view everyday psychology as somehow incommensurable with science, as offering a quite different kind of understanding from science. And other philosophers, while allowing that folk psychology may have pretensions to science, hold that it fails miserably to live up to them.

I intend to by-pass this issue in most of the rest of this book, by stipulating that, unless I say otherwise, my use of folk psychological talk is to be understood as a place-holder for the true special science of psychology. So philosophers who think that folk psychology is already a science can take my words at face value. On the other hand, those who think something different is needed for a genuine cognitive science should simply understand my psychological talk as referring indirectly to their own favoured cognitive states. There remain the pessimists who think that cognitive science of any kind is impossible, that there cannot be a theory of our cognitive workings that stands to our physics and physiology as meteorology, say, stands to the physics of the atmosphere. To these pessimists I simply concede that if their bet about the future of cognitive science is right, then a number of the issues I address in this book do not arise. (Though in fact the issue of this chapter, the relationship between the psychological and the physical3, arises not only for optimists who accept the possibility of a high-level psychological science, but also for those pessimists of a Wittgenstenian or Davidsonian bent who reject this possibility but nevertheless uphold everyday psychology as a respectable but non-scientific form of discourse. For they too need to consider the relationship between psychological states and brain states. It is only pessimists who take the eliminativist line and reject high-level psychological thinking of any kind who can avoid addressing the mind-body problem.)

My own view, for what it is worth, is that everyday psychology constitutes an impressive theory from a scientific point of view, capable of improvement and refinement, of course, and with a number of philosophically puzzling aspects, but certainly containing a great deal of predictive information, and quite probably giving

some insight into the structure of our internal workings. I prefer to avoid, however, debates about whether its undoubted imperfections merely mean it is a somewhat inaccurate theory about real entities (like nineteenth-century atomic theory), or whether they make it a false theory about imaginary entities (like the eighteenth-century caloric theory of heat). This issue would be hard enough to resolve if we knew the whole psychological truth (though then it wouldn't matter very much). But, as it is, there are better things to think about.4

1.3.3 Among the ways in which psychology is philosophically puzzling is that it deals in propositional attitudes: its explanations invoke beliefs, desires and other states which represent things as being a certain way. (And we can expect the states of any future cognitive science to be similarly representational.) In chapter 3 below I shall address the topic of mental representation. At this stage we need only note that representation complicates the issue of the supervenience of the mental on the physical. For, as a number of writers have observed5, there are many plausible cases of two people having physically identical brains, and yet having propositional attitudes with different representational contents. These examples imply that psychological states individuated by representational content don't supervene on the physics of the brain.

Physicalists about psychology have two options here. They can argue that any such "broad" psychological state is really a kind of relational state, and that therefore, in the way indicated earlier in this chapter, physicalism only requires the supervenience of such states on the physical properties of some larger system which includes the individual's brain as a part. Alternatively, they can argue that such broad states are not really part of serious psychological theory, and therefore that their non-supervenience is not a problem for physicalism about serious psychology. In what follows I shall defend the former line, in particular in sections 1.5 and 1.7 below.

1.4 The Completeness of Physics

Now for the arguments in favour of physicalism. In what follows I shall consider two different such arguments. But both arguments will hinge on what I shall call "the completeness of physics". So in this section let me explain what I mean by this.

I take it that physics, unlike the other special sciences, is complete, in the sense that all physical events are either determined, or have their chances determined, by prior physical events according to physical laws. In other words, we need never look beyond the realm of the physical in order to identify a set of antecedents which fixes the chances of any subsequent physical occurrence. A purely physical specification, plus physical laws, will always suffice to tell us what is physically going to happen, insofar as that can be foretold at all.

Note that not all subject areas are complete in this way. For instance, meteorology is not complete. Some weather phenomena arise from antecedents which are not themselves weather phenomena. The beat of a butterfly's wing, students of chaos tell us, can play a part in determining next week's cyclone. Less exotically, psychology is obviously not complete, given that plenty of mental events result from non-mental ones, as when I sit on a drawing pin and feel a pain. But physics is special in this

respect. If we take any physical result, and look back in time to see what gave rise to it, then, I say, prior physical factors will always suffice to give us as full an explanation of that result as is possible.6

I have stated the the completeness of physics baldly, as something to which all will assent. But perhaps some readers will have doubts. I can imagine two possible sorts of worry. The first would be a general worry that the completeness of physics is an empirical claim and therefore inadmissible in a philosophical argument. I have nothing to say to this beyond the points made earlier in the introduction to this book. The second worry would be more specific: even if empirical claims are admissible in philosophy, is the completeness of physics really a well-supported empirical claim? In particular, what exactly does "physics" mean here? On some perfectly natural ways of reading this term, the completeness of "physics" seems false.7 However, let me postpone discussion of this second worry to section 1.9 below. For the moment it will be helpful take the completeness of physics at face value and see what would follow if it were true. We will be better placed to evaluate queries about it when we see how it matters to physicalism.8

1.5 The Manifestability Argument for Supervenience

Consider now the following argument for the supervenience of psychology on physics.

Premise (1). According to the completeness of physics, the chances of physical consequences are fixed, once physical antecedents are given. So if two systems are physically identical and in the same physical contexts, they will issue in the same physical consequences or chances thereof.

Premise (2). Now add in the assumption, which I shall call the "manifestability of the mental", that if two systems are mentally different, then there must be some physical contexts in which this difference will display itself in differential physical consequences, or at least in differential chances for such consequences.

Conclusion. It follows that mental differences without physical differences are impossible. (1) tells us that physical identity guarantees identity of physical consequences or chances thereof. And (2) tells us that mental difference requires the possibility of different physical consequences or chances thereof. So physical identity rules out mental difference.9

The crucial idea here is that the completeness of physics leaves no room for mental differences, or any other differences, to make a difference to physical consequences, once physical antecendents are given. Physical categories by themselves always suffice to fix the chances of physical consequences, without the help of mental categories. So the only way for mental differences to be manifestable is for them to have different physical bases.10

The two premises to this argument are the completeness of physics and the manifestability of the mental. As I said, I shall come back to the completeness of physics at the end of the chapter. Here we need to consider the manifestability of the mental. The most obvious argument for this principle would be that mental

differences must always be capable of showing themselves in differential behaviour: there certainly seems something initially odd about the idea of two people who are mentally different, yet behave in the same way in all physical contexts. (In this connection, note that the manifestability principle is not the strong requirement that every particular mental difference actually manifests itself in differential physical consequences; just the weaker assumption that, for any type of mental difference, there is some type of physical context in which that difference would be physically manifested.)

If this behavioural interpretation of the manifestability principle were acceptable, then a strong version of the supervenience of the psychological on the physical would follow, namely, the supervenience of psychological states on brain states. For we could run the argument as follows. Mental difference require behavioural differences. But behavioural difference are fixed specifically by prior brain states. So there can't be mental differences without brain state differences.11

However, there are good reasons for denying that all psychological states supervene on brain states I am thinking here of the kind of "broad" propositional attitudes mentioned in section 1.3.3. As we saw, the distinguishing characteristic of broad attitudes is that individuals with identical brains can fail to share them. So it follows from the argument in the last paragraph that a manifestability requirement in terms of behavioural displays is too strong a requirement for broad attitudes. And this is of course what we do find: differences in in broad attitude don't automatically display themselves in behavioural differences. To take a familar example, consider Carl, who wants a glass of H2O, and Lrac, his physically identical Twin Earth counterpart, who wants a glass of XYZ. They have different broad attitudes. But their behaviour, in the sense of the physical movements of their bodies, will be the same in all physical contexts.

As I observed in 1.3.3, the failure of broad attitudes to supervene on brain states does not mean that physicalism is false. For if broad states are relational states, then it will suffice for physicalism that they supervene on the physics of the individual-system-and-relevantly-related-systems, even if not on the physics of the individual system alone. So it remains possible that the general manifestability argument for supervenience might still establish this weaker kind of supervenience for broad beliefs, even if not supervenience on brain states. All we need is a weaker manifestability premise to the effect that differences in broad beliefs are somewhere manifested in physical consequences, even if they are not manifested in behavioural consequences.

In defence of this weaker version of the manifestability premise, note that a mental difference which was not physically manifestable in any way would be radically undetectable. We know that our sense organs work by physical interaction with the environment, as do the instruments and other aids by which we extend the power of our sense organs. So if two different mental states yielded exactly the same physical manifestations in all contexts, then there would seem no possibility of our ever finding out about their difference. Yet surely any real mental difference ought to be somehow detectable, even if not behaviourally.

To illustrate this point, note that even the broad mental difference between Carl and his identical doppelganger Lrac will be distinguished by some differential physical

consequences. For this broad mental difference depends on the relational difference that, where Carl is surrounded by H2O, Lrac is surrounded by XYZ. And this difference in their environments will obviously produce some differential physical consequences by which we can distinguish the two cases.

I recognize that this defence of the manifestability requirement, and hence of supervenience, is less than fully principled. For one thing, it leaves it open for opponents of physicalism to object that it is possible that there be mental differences that are not in any way detectable by human beings. More pressingly, opponents of physicalism could also query whether our ability to detect mental differences always depends on physical interaction with our environments. Thus anti-physicalists might argue that our access to conscious mental states in particular is primarily via introspection, rather than via the normal five senses, and that there is no immediate reason to suppose that the deliverances of introspection are mediated by physical processes, however it may be with the other senses. This would then open the way for anti-physicalists to argue that conscious states might fail to supervene on physical goings-on, and so that conscious differences need not manifest themselves physically, and yet to hold that those differences could still be detectable, via introspection: for example, they could argue that you could be in just the same physical state at two different times, and yet know introspectively that you were in pain at one time and not at the other.

I shall not attempt to plug this particular gap in the manifestability argument, however. For there is a rather more basic flaw in the argument, to which I shall now turn. To deal with this more basic flaw, we will need to switch to a significantly different form of argument for physicalism. Moreover, this alternative form of argument will be immune to the anti-physicalist appeal to non-physical introspective powers.

1.6 Manisfestabilty is Not Enough

To understand the more basic flaw in the manifestability argument, recall how I earlier alluded to the possibility of different ways of understanding supervenience, depending on how strongly we read the "determine" in "physical features determine special features" (or, equivalently, on how strongly we read the "cannot" in "cannot differ in special features without differing physically").

A weak version of supervenience would understand these notions in terms of natural necessity: that is, it would take physical features to determine special features across all possibilities where we hold the actual laws of nature fixed; equivalently, it would say that any two systems governed by the actual laws of nature cannot be different in any special respects without differing physically.

A strong version of supervenience would do it in terms of "metaphysical" necessity rather than natural necessity: supervenience requires that physical nature determines special nature across all possible worlds whatsoever; no two possible systems of any kind can be different in some special respect without being physically different.

Now only the stronger of these versions of supervenience constitutes a plausible explication of physicalism. To see why, we need only consider epiphenomenalism, the doctrine in the philosophy of mind which holds that mental states "float above" the brain as distinct conscious phenomena, not responsible for any physical effects themselves, but nevertheless causally determined by the physics of the brain, and so incapable of varying without physical variation. Epiphenomenalism implies supervenience in the weak sense, since it implies that, if we hold all natural laws constant (including in particular the putative epiphenomenalist laws by which by physical brain states cause conscious states) then physical nature will determine mental nature: identical physical brain states, plus the laws according to which physical brain states cause conscious states, will ensure identical conscious states. But epiphenomenalism clearly isn't a physicalist doctrine, since it explicitly specifies that conscious mental properties are ontologically distinct from physical ones. So weak supervenience clearly does not suffice for physicalism.

However, the same objection does not apply if we equate physicalism with strong supervenience. For while epiphenomenalism does imply that the mental is fixed by the physical across all natural pssibilities, it does not imply that such brain-mind determination holds across all possibilities, including those possible scenarios where these epiphenomenalist brain-mind laws break down. This is because epiphenomenalism insists that conscious properties are ontologically distinct from physical ones. So it implies that it is metaphysically possible, even if not "naturally" possible, for a creature physically just like me, say, to have different conscious states, or indeed to have no conscious states at all. While such a creature would violate epiphenomenalism's putative natural laws of mind-brain causation, and so fail to be "naturally" possible, epiphenomenalism allows that these laws themselves are not absolutely necessary, and so implies that such a creature is metaphysically possible. Conversely, the doctrine that such a creature is not metaphysically possible would be inconsistent with epiphenomenalism's distinction of conscious mind from physical brain, and so would constitute a plausible explication of genuine physicalism.

After all, strong supervenience says that it is metaphysically quite impossible for two beings to differ in some special property unless they differ physically. But how could this be so absolutely impossible, unless the special property was itself in some sense itself physical? If the special property weren't itself physical, then surely there would be metaphysical room, if not natural room, for it to float free of the physical realm, in the way the epiphenomenalist's conscious properties float free of physical properties in worlds with different brain-mind causal laws. So it looks as if strong supervenienceÑthat is, the denial of any metaphysical room for special properties to float free of physical onesÑwill indeed ensure that special properties are physical.12

To return to the original issue of this section, the basic flaw in the manifestability argument for physicalism is that it only constitues a good argument for weak supervenience, not for strong supervenience, and so fails to establish physicalism. The fault lies with the manifestability premise, that is, the premise according to which mental differences must manifest themselves in differential physical consequences. For any version of this premise strong enough to deliver genuine physicalism would blatantly beg the question against non-physicalist views like epiphenomenalism.

Consider what epiphenomenalists would say about the manifestabilty premise. They would happily allow that mental differences will display themselves in differential physical consequences as long as the laws by which brain states cause conscious states are held constant: given these laws, then different conscious states must have been caused by different physical states, and we can expect these physical differences to have different physical consequences. But epiphenomenalists will point out that there is no need to expect this manifestability requirement to hold up across all possible worlds, including worlds where the actual brain-mind laws break down. After all, if we allow, as epiphenomenalists will, that there are metaphysically possible worlds in which I have physical duplicates with different conscious states, or with no conscious states at all, then we will not expect these mental differences, between me and my other-minded physical duplicates, to display themselves in any differential physical consequences.

So epiphenomenalist anti-physicalists will see no reason to concede that the manifestability premise holds across all metaphysically possible worlds, even if it holds in all naturally possible worlds. And correspondingly, they will not view the manifestability argument as providing any substantial reason to suppose that the mental supervenes on the physical across all metaphysically possible worlds. They can allow that mental differences will display themselves in differential physical consequences as long as we hold all laws of nature fixed, and correspondingly concede the weak supervenience thesis that mental differences without physical differences are naturally impossible. However this, as we have seen, falls short of physicalism proper. Genuine physicalism requires strong supervenienceÑmental differences without physical differences are metaphysically impossible. But epiphenomenalists will see nothing in the manifestability argument to force them to this stronger claim, for they will have no inclination to accept that all metaphysically possible mental differences must display themselves in differential physical consequence.13

1.7 The Causal Argument for Physicalism

Let me now turn to a somewhat different argument for physicalism, which I shall call "the causal argument". This argument, like the manisfestability argument, will hinge on the completeness of physics. But instead of appealing to requirements on the manifestation of mental states, it will appeal to the possession of causal powers by mental states. This shift of focus will yield a more effective line of reasoning against anti-physicalist view like epiphenomenalism.

Thus consider the following premise, which I shall call the "principle of mental efficacy":

Premise (3). Every mental occurrence causes some physical effect.

Note now that, on just about any account of causation14, the following is an immediate corollary of the completeness of physics:

Premise (4). All physical effects have complete physical causes ("complete" in the sense that those causes on their own suffice by physical law to fix the chances of those effect).

Consider now some mental occurrence, and one of the physical effects which are required by (3). For example, suppose you decide to lift your arm, and as a result your arm rises15. By (4) this physical effect will also have a complete physical cause, which will presumably involve the neuronal and other physical antecedents of your arm rising. So it follows that your arm rising has two causes: a mental cause, your decision, and also a physical cause, your neurones firing.

Does this mean that such physical effects are always overdetermined, like the death of the man who is shot and simultaneously struck by a random bolt of lightning? This doesn't seem right. After all, when an effect is overdetermined by two causes, it follows that it would still have occurred if either one of the causes had been absent: the man would still have been killed by the lightning bolt even if he hadn't been shot, and vice versa. But we don't similarly want to say that your arm would still have gone up even if you hadn't wanted to lift it, or, alternatively, even if different neurones had fired in your brain.

The obvious conclusion is that your desire and your neurones are not two independent causes, like the shot and the lightning bolt, but are in some sense the same cause. We need somehow to identify the mental cause with the physical cause, so as to avoid the conclusion that the movement of your arm was overdetermined16.

Note how this argument differs from that in the last section. There the aim was to show that the physical always co-varies with the mental, and the argument was that physical variation is needed to produce the external evidence for mental variation; the trouble was that this argument only established co-variation across naturally possible worlds, which was too weak for physicalism. In this section the aim has been to show that the mental is ontologically inseparable from the physical, and the argument has been that such a separation would imply an absurd proliferation of causal overdetermination; if this ontological inseparability does follow, it will mean that there is no metaphysical room for mental properties to float free of physical ones, and so will establish genuine physicalism.

It might seem as if the causal argument begs the question against anti-physicalist epiphenomenalism just as much as the manifestability argument. Epiphenomenalists, after all, will deny the assumption of causal efficacy, just as they denied any strong manifestability premise. So they will escape the causal argument too. They don't need to explain why bodily movements aren't always overdetermined, since they don't admit they have mental causes in the first place.

But there is a significant dialectical difference between the two cases. There is nothing pre-theoretically objectionable about the denial of a strong manifestability premise: nothing obvious will go wrong with our overall view of the world just because we allow the mere metaphysical possibility of mental differences without physical manifestations. By contrast, it clearly flies in the face of any number of normal assumptions to deny that mental events have physical effects. If my conscious thirst isn't what causes me go to the fridge for a beer, and my conscious map-reading

isn't what causes me to choose one route rather than another in a strange city, and so on, then we are going to have to think again about most of our assumptions about the way the human world works.

Given this, we can well ask why the epiphenomenalist wants to adopt the curious view that conscious mental states are causally inefficacious, especially given the availability of physicalist alternatives which avoid it. The only plausible answer, I take it, is to do with consciousness: epiphenomenalists are persuaded that any physicalist account of the mental will leave out the essential conscious features of the mental, and so are persuaded to postulate a distinct, non-physical realm of mental events, even at the cost of denying that the mental affects the physical. I shall return to this issue in chapter 4 below, where I shall argue that there is nothing in consciousness that is left out by physicalism, and therefore that the epiphenomenalist denial of the causal efficacy of the mental is ill-motivated.

Before proceding, let me quickly deal with one complication. This relates to broad mental states. We saw earlier how broad mental states complicated the manifestability argument. Similar complexities arise in connection with the causal argument.

Thus note how I illustrated the causal argument by focusing on the bodily effects of mental states, like arms rising, and then inferred that the mental causes of these bodily effects must be identical with their "neuronal causes". However, this specifically neuronal conclusion sits ill with the possibility of broad mental states. For broad mental states can't be identical with internal brain states, given that they depend on matters outside the skin. Carl and Lrac differ in their respective desires for H2O and XYZ, even though they are internally physically identical. In line with this, it seems wrong to say that their different desires cause their bodily movements: bodily movements are surely caused by matters inside the skin, not by features that stretch outside.

Still, the fact that broad mental causes can't be the same as brain states doesn't mean they can't be equated with any physical states, in particular with certain physical features of their possessors-and-relevantly-related-systems. And it is not hard to see how the causal argument might be made to deliver this weaker conclusion. All we need is a causal efficacy premise to the effect that broad mental states cause some "broad" physical consequences, even if they don't cause the bodily movements that result from neuronal causes alone. And there seems no difficulty about this version of the efficacy premise. For example, Carl's desire may cause a glass of H2O to move, where Lrac's desire will cause a glass of XYZ to move. And then, with the efficacy premise so restored, we can use the causal argument to argue that Carl and Lrac's desires must be equated with those physical feaures of themselves-and-their-surrounding-environments which are responsible for these broad effects.

1.8 Generous Causation and Alternatives to Type Identity

I argued in the last section that the mental causes of physical effects must be the same as the physical causes of those effects. Exactly how we construe this equivalence, however, depends on what view we take of the ontological status of causes in general.

Some philosophers, most prominently Donald Davidson (1967), think that causation is a relation betwen events construed as "bare particulars" shorn of any general attributes. However, there are good arguments for being dissatisfied with this anaemic view of causation, and for preferring to view causal relata as facts rather than as Davidsonian bare particulars17. Accordingly, I shall assume the factual view of causal relata in what follows.

However, if you view causation as a relation between facts in this way, then it may seem as if the causal argument is in danger of proving too much. In particular, it may seem in danger of proving that mental properties must be type identical with physical properties, notwithstanding the intrinsic implausibility of this type identity claim. For, if causes are facts, then the causal argument's conclusion, that mental causes must be identical with physical causes, will require that mental factsÑsuch as that I am in pain, sayÑare identical to certain physical factsÑI have a certain brain feature, sayÑand these two facts cannot be identical unless the properties they involveÑbeing in pain, having that brain featureÑare themselves identical.

Well, this type identity would indeed follow from the causal argument if we take a very strict view of causation, and insist that the only thing that can cause a physical effect is another strictly physical fact. For then the principle of mental efficacy, according to which mental facts cause physical effects, can only be satisfied if mental facts are themselves instantiations of strictly physical properties. However, suppose we understand causation in a more generous sense, and allow that an instance of a strongly supervening property causes the effects of those facts on which it supervenes. Then the principle of mental efficacy will only require that mental properties are type identical to physical properties or that they strongly supervene on physical properties. For as long as the latter possibility is realized, then it will still be true, in the generous sense, that mental facts cause the physical effects of the physical facts on which they supervene.

As an illustration of this possibility, consider the functionalist view that mental states are causal intermediaries between perceptual inputs and behavioural outputs. The orthodox version of this view does not identify pain, say, with whichever first-order property mediates causally between damage detectors and avoidance behaviour in any given species. For this would have the "chauvinist" implication that species with different internal workings could not share the experience of pain. Rather the standard functionalist view is that pain is a second-order property, the property-of-having-some-property which mediates causally between damage detection and avoidance behaviour, which second-order property can therefore be present across beings with different innards.

Now, on this functionalist view, pains can't cause bodily movements in the strict sense which requires identity with strictly physical facts. For, if pains are instantiations of second-order properties, they cannot be identical with any first-order physical facts. Still, such functionally understood pains can still be "realized" by physical properties, in that they can be present purely because some first-order physical fact which mediates between damage and avoidance is presentÑand in that case a pain will indeed cause bodily movements, in the generous sense in which superveners cause what their subveners cause. For if a pain is so realized by a physical fact, then it will supervene on this physical fact, even though not identical

with it, in that any metaphysically possible being with this physical property will be in pain, since it will possess the property-of-having-some-property which mediates causally between damage and avoidance.

Let me clarify my direction of argument here. I am not at the moment concerned to uphold functionalism, nor, consequently, am I particularly concerned to argue that functionalist mental definitions are satisfied by physical states in humans18. Rather I have introduced functionalism merely as an illustration of how facts that are not themselves physical facts can nevertheless cause physical effects, at least in the generous sense of causation.

More generally, if we understand the causal argument in terms of the generous sense of causation, then the conclusion will be mental facts must in some way strongly supervene on physical facts (otherwise mental facts couldn't cause physical facts even in the generous sense, given the completeness of physics). Functionalism offers one illustration of how this might be so, even when type identities are not available. But my conclusion is not that functionalism must be true, only that the mental must somehow strongly supevene on the physical.

For a further example of a theory of this form, consider Donald Davidson's view of the mental. (Davidson, 1980, passim. Though Davidson's view of the mental is standardly presented in harness with the Davidsonian view of causation mentioned above, it is helpful to separate out these two aspects of Davidson's thinking.) Davidson holds, in effect, that to be in a given mental state M is to be in some state which causes behaviour which would warrant the attribution of M to you. This is a different theory from functionalism, since it makes essential appeal to the non-scientific canons of interpretation which Davidson takes to govern our attributions of mental states to others. But, just like functionalism, it allows room for the idea that the mental may be realized by the physical, and consequently strongly supervene on it. For if it is physical state P which causes the behaviour which warrants the attribution of mental state M to person X, then X will be M purely in virtue of being in P, and correspondingly any possible creature with P will have M, since it will have some state which causes behaviour which would warrant the attribution of M.

So the Davidsonian view, like functionalism, will also satsify the requirement that mental facts should cause physical facts, at least in the generous sense. Still, as with functionalism, I mention this, not as an argument for the Davidsonian view in particular, but simply as another illustration of how the requirement of supervenience on physical states allows the causation of physical effects, even in the absence of type identity.

I have no clear views about the full range of ways in which mental properties might supervene on physical ones. Functionalism and Davidsonianism are two such options, but there may well be others. However, there is no need to decide this issue here. It will be enough if I have shown that some such view of the mental is demanded by the causal argument.

Of course, there remains the option of embracing epiphenomenalism, and denying that the mental is efficacious, even in a generous sense. As I observed earlier, the normal motivation for this unpalatable view is to do with consciousness, and the

conviction that conscious states at least must be ontologically quite distinct from any physical states. The question this raises, and to which I said I shall return in chapter 4, is whether this anti-physicalist convinction about consciousness rests on solid enough grounds to justify the radical step of denying that our thoughts and feelings affect our actions. But for the moment I am content merely to point out that the minimum price for rejecting physicalism is epiphenomenalism. Anti-physicalists need to deny some premise in the causal argument, and the line of least resistance is to deny the principle of mental efficacy.

1.9 The Completeness of Physics Defended

There is an alternative, if less obvious, way to resist the causal argumentÑnamely, by denying the completeness of physics. This assumption may seem initially plausible. But, as I allowed earlier, it is by no means entirely unproblematic.

The central difficulty facing defenders of this assumption is an obvious dilemma about what they mean by "physics". Either "physics" means the theory currently taught in university departments of physics and presupposed by articles in physics journals, or it means some ideal future theory that will succeed current theory.

The trouble with the first horn of this dilemma is that, if the past form of physical theorizing is anything to go by, current physics is no doubt inadequate in certain respects, and in particular in failing to identify all the antecedents for certain physical effects. So current physics is not complete.

The trouble with the other horn, by contrast, is that we don't yet know what physical categories will be assumed by the ideal future physics. So we scarcely seem to be in any position to maintain that those categories will suffice for complete explanations of all physical effects.

However, I think there is a version of the second horn of this dilemma which will serve the purposes of the arguments of this chapter. Suppose we simply define "physics" as the science of whatever categories are needed to give full explanations for all physical effects. I accept, as above, that this science will be different from current physical theory, and thus that we don't yet know what it is. But, even so, there is no difficulty about how we know that it is complete, for we have simply defined it so as to be complete.

The obvious worry about this definitional strategy is that it seems to remove any significant content from the thesis of completeness, and thereby to make it doubtful that the thesis could have any substantial conclusions. There are two dimensions to this worry. First, the definitional strategy characterizes physics as the science of whatever is needed to explain "physical" effects. But what are "physical effects", if we haven't yet specified what counts as "physics"? Second, even if we had some independent hold on "physical effects", the proposed strategy would still make the completeness thesis an empty analyticity, for it simply defines "physical" categories as all those needed to explain physical effects, from which completeness immediately follows.

Let me deal with these two worries in turn. To deal with the first worry, I propose that we simply postulate some pre-theoretically given class of paradigmatic physical effects, such as stones falling, the matter in our arms moving, and so on. If we take this class to be independently given, then we can effectively characterize the rest of physics as all the categories that need to be brought in to explain those paradigmatic physical effects.

But this still leaves us with the second worry, that even we help ourselves to a pre-theoretical class of paradigmatic physical effects, we are still defining physics in such a way as to make the completeness of physics a matter of definition. I still need to explain how substantial conclusions about the truth of physicalism could possibly follow from such a definition19.

My answer is that no substantial conclusions follow from the completeness of physics per se. But they do follow from the joint assumption that (a) physics is complete and (b) that it does not make any use of psychological categories.

Let me explain. In itself, the above definition of physics leaves it open that psychological categories may turn out to be needed as an essential part of physics. Maybe psychokinesis is true, and there are physical effects that can't be accounted for without making essential mention of distant volitions. Less exotically, maybe some bits of behaviour can't in fact fully be accounted for purely in terms of muscular activation, neuronal activity, and so, without bringing in extra mention of prior mental states. Now, if psychological categories do turn out to be needed to give full explanations for physical effects in this way, then the issue of whether psychology supervenes on the physical, as I have defined it, becomes trivial. Psychology will indeed supervene on the physical, but only because it is included in the physical, not because psychological variation requires variation in something else.

On the other hand, if psychology is not part of the physical, as I have defined it, then the arguments of this chapter will go through as before. That is, if psychological categories are not in fact ever essential to explaining physical effects, then physics, in the sense of whatever is needed to explain physical effects, will be both complete and exclusive of psychology, and the arguments of this chapter will show that psychological states are non-trivially supervenient on physical states.

It seems to me highly unlikely that the psychological will turn out to be part of the physical. Current physics, I take it, aims to develop a complete theory of paradigm physical effects in terms of the categories of energy, field and spacetime structure. I am quite prepared to believe that this this aim cannot be achieved, and that the categories of current physics will need supplementation before we can get a genuinely complete theory. What I do not believe is that they will need supplementation by psychological categories.

I am here making an empirical claim. The history of science yields a great deal of empirical evidence about the kind of causes that are responsible for the motion of stones and other kinds of matter. This evidence does not, perhaps, allow us to formulate a definitive list of all the necessary categories. But it does, it seems to me, provide sufficient grounds for concluding that mental categories are not among them.20

To help see what is at issue here, it is illuminating to consider Descartes' views on the matter. Descartes did think that there were physical effects that could not be explained without bringing in mental antecedents. Descartes believed that the total amount of motion, in the sense of mass times speed, is conserved, according to regular laws, in all material interactions, and therefore that the speeds of all material bodies are determined by earlier such speeds. However, unlike us, Descartes did not believe in the conservation of momentum, considered as a directional quantity, and so did not think that the direction of motion of material bodies was necessarily determined by prior physical factors. And it was this gap that Descartes exploited to explain how the mental, although ontologically quite distinct from the physical, can nevertheless affect the physical: the mental interacts with the physical in the pineal gland, and influences the direction of motion of certain particles (though not their speed, since this is always fixed by prior physical states).

To hold that the psychological is part of the physical is to believe a version of what Descartes believed, namely, that there is a gap in the determination of certain physical effects, which can only be filled by mental occurrences. And this is what seems highly unlikely to me. It is one thing to hold that the current categories of energy, field and spacetime structure leave a gap in the determination of certain physical effects. It is another to hold that this gap cannot be filled without bringing in the mental. If that were true, after all, then the obvious moral would be that physicists needn't build expensive particle accelerators to generate theoretically anomalous physical phenomena; instead they could find plenty of currently inexplicable physical phenomena simply by looking inside people's heads.21 I think we have good empirical reason to reject this possibility as absurd.

Extra References

Papineau, D. 1996: "Theory-Dependent Terms" Philosophy of Science 63

[Because of this new reference, remove Stich (1991) from the Bibliography.]

Steward, H. 1996. "Comments on Philosophical Naturalism" Philosophy and Phenomenological Research 57

Witmer, G. 1998. "What is Wrong with the Manifestability Argument for Supervenience" Australasian Journal of Philosophy 76

1 For arguments in favour of physicalism, see Lewis (1966), Davidson (1970), Peacocke (1979, ch III.3), McGinn (1982, p 29), Smith and Jones (1986, pp 57-59), McFetridge (1990, p 86), Lycan (1987, pp 2-3). Reasoned opposition to physicalism is offered in Crane and Mellor (1990), Crane (1991). Most of these contributions will be referred to further in what follows.

2 It is sometimes suggested that this kind of shift, from a "local" to a more "global" supervenience, makes room for ad hoc defences of supervenience, and so dilutes physicalism beyond interest. To answer this charge, we should require that wider systems should be admitted as subvening bases only if there are independent grounds,

apart from a desire to save physicalism, for regarding the putatively supervening properties as relational.

3 For brevity I shall often focus in this chapter on the relation between the psychological and the physical. But the analysis will be of general significance, as the structure of my arguments will indicate.

4 I think that there is good reason to think that theoretical concepts in general, and psychological concepts in particular, are vague, in that there will often be no fact of the matter about how to apply them to cases where the theory that defines them breaks down. For more on this see Papineau (1996).

5 See Putnam (1975), Burge (1979, 1982) and Evans (1982).

6 Some readers might baulk at my use of "explanation" here, on the grounds that a full physical specification of the antecedents of some large-scale physical outcome won't necessarily be illuminating for us humans, in the way that an explanation using chemical or biological or psychological terminology might be (cf Putnam, 1978, p 42). No matter. My argument only requires that the physical antecedents fix or cause the physical outcome, not that they illuminate it. David Owens (1992) is even more particular, and would baulk at this last use of "cause", on the grounds that causes aren't causes unless they illuminiate. Again no matter. My arguments need only whatever is left in the notion of cause after we take away the anthropocentric factor of illumination.

7 Cf Crane (1991).

8 As it happens, when I do return to the completeness of physics in 1.9, I suggest that this thesis itself can most usefully be understood as an analytic truth, rather than as an empirical claim. However, when we do read it in this way, the burden of my argument for physicalism is then taken up by some closely related empirical assumptions.

9 This argument is found in McGinn (1982, p 29) and further discussed by McFetridge (1990, p 86). In Papineau (1990) I tried to run the argument with a weaker version of premise (2), requiring only that mental differences have some different consequences, not necessarily physical ones. But when I presented this version of the argument at the Analysis 50 Conference in Cambridge, Tomis Kapitan showed me that it begged some crucial questions.

10 Why doesn't the argument work in reverse, and also show that all physical differences must depend on mental differences? The essential reason is that the mental is not complete. Even if we accept, as is not entirely implausible, the "mental manifestability of the physical" ("if two systems are phyiscally different, there must be contexts in which this will produce differential mental effects"), we cannot conclude that these differential mental effects must always depend on prior mental differences, for lack of the premise that mental effects are always fixed by mental antecedents.

11 This is the version of the argument articulated by McGinn. He does, however, observe that it may not apply to all mental states.

12 Some readers may be wondering whether this equation of physicalism with strong supervenience has not simply taken us a long way round back to the earlier equation of physicalism with type identity. For haven't I just argued that the virtue of strong supervenience is precisely that it ensures that special properties are ontologically inseparable from physical properties, by contrast with weak supervenience, which only requires that special properties are correlated with physical properties by the actual laws of nature, but need not be ontologically intertwined with them? Well, the virtue of strong supervenience is indeed that it ensures an ontological dependence of special properties on physical properties, and not just a correlation. But the point of formulating physicalism in terms of supervenience, rather than type identity, is precisely that it is possible to have such ontological dependence even when type identities are not available. I shall return to this in section 1.8 below.

13 For further discussion of the failings of the manifestability argument, and for other criticisms of the original English version of this chapter, see Steward (1996) and Witmer (1998).

14 David Owens (1992) is an exception. But, as I said in footnote 6, I could grant Owens his stronger notion of cause, and simply phrase my arguments in terms of a weaker one.

15 Are bodily movements, like arms raising, mouths moving, and so on, properly counted as physical effects? Strictly, no. "Arm" and "mouth" are biological terms, not physical ones, and it is doubtful that they can be reduced to physical notions. So for full accuracy we ought to take the physical effects of mental causes to be the motion of bits of matter, which happen to be in arms, mouths, and so on. However, it will smooth the exposition if I can be less than strict on this point.

16 This form of argument for token congruence is to be found in Peacocke (1979, ch. III.3). It has obvious affinities with the discussion in Davidson (1970).

17 For a defence of this view. see Mellor (1987). Another alternative to Davidson's view of causation is to allow that causes are events, but insist that events are instantiations of properties, rather than Davidsonian bare particulars (Kim, 1973). However, Mellor (op cit)argues that "events" of this kind are simply a subspecies of facts.

18 David Lewis (1966) does argue from a version of functionalism to mind-brain identity. Lewis's argument shares one premise with my causal argument, namely, the completeness of physics. But where my other premise is only that each particular mental cause has some physical effect, Lewis makes the stronger functionalist assumption that different mental types can distinguished by their characteristic causal role in mediating between physical causes and effects. (He then concludes, from the completeness of physics, that such roles are always filled by physical states.)

19 Crane (1991) argues on just these grounds that the version of my argument for physicalism in (Papineau, 1990) collapses into triviality.

20 Let me guard against one possible source of confusion here. When I say that a compete physics excludes psychology, and that psychological antecedents are therefore never needed to explain physical effects, the emphasis here is on "needed". I am quite happy to allow that psychological categories can be used to explain physical effects, as when I tell you that my arm rose because I wanted to lift it. My claim is only that in all such cases an alternative specification of a sufficient antecedent, which does not mention psychological categories, will also be available. I need the thesis that psychological terms are not included in the minimal set which provides sufficient conditions for all physical effects, not that they are not included in any such set.

21 Cf. Lycan (1987, pp 2-3).