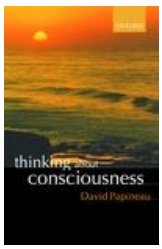

Thinking about Consciousness

David Papineau

https://doi.org/10.1093/0199243824.001.0001

Published: 2002 **Online ISBN:** 9780191598166 **Print ISBN:** 9780199243822

CHAPTER

1 The Case for Materialism

p. 13

David Papineau

https://doi.org/10.1093/0199243824.003.0002 Pages 13–46

Published: April 2002

Abstract

The basic causal argument for materialism is that since physics is causally complete, conscious states must either be physical, or they must be epiphenomenal "danglers" with no causal influence on the physical world. Papineau explores this argument in detail, paying particular attention to the concept of causation, the meaning of "physical", the relevance of functionalism and other versions of nonreductive physicalism, and the status of the claim that physics is causally closed.

Keywords: causal argument, causal closure of physics, causal completeness of physics, causation, consciousness, functionalism, matter, nonreductive physicalism

Subject: Epistemology, Philosophy of Mind

1.1 Introduction

Books on consciousness often begin by distinguishing between different kinds of consciousness. We are told about self-consciousness and sentience, creature consciousness and state consciousness, phenomenal consciousness and access consciousness, perceptual consciousness, higher-order consciousness, and so on. I'd rather leave all this until later. Some of these distinctions will become significant in due course, and will be explained when they are needed. Others will not matter to my discussion.

For the moment, all I want to say is that I am concerned with that aspect of consciousness that makes it so philosophically interesting. Namely, that having a conscious experience is *like something*, in Thomas Nagel's striking phrase (1974). It has become standard to use 'phenomenal' or 'subjective' to focus on this feature of consciousness, and I shall adopt these usages in what follows.

The idea is best introduced by examples rather than definitions. ('If you gotta ask, you're never gonna know.') Compare the difference between having your eyes shut and having them open, or between having your teeth drilled with and without an anaesthetic. When your eyes are open, you have a conscious visual
p. 14 experience, and ↳ when your teeth are drilled without an anaesthetic, you have a conscious pain. It is like something for you to have these experiences. It is not like that when you close your eyes, or when the

anaesthetic takes effect. What you lose in these latter cases are elements of phenomenal or subjective consciousness.[1] From now on, when I say 'conscious', I shall mean this kind of consciousness.

Much of what follows will be concerned with a particular philosophical puzzle about consciousness: namely, the puzzle of how consciousness relates to the physical world. There are other philosophical puzzles about consciousness, but this seems to me the most immediate. We will be ill placed to understand anything about consciousness if we cannot understand its relation to the physical realm.

The puzzle can be posed simply. On the one hand, there is a strong argument for adopting a materialist view of conscious states, for supposing that conscious states must be *part* of the physical world, that they must be *identical* to brain states, or something similar. Yet, on the other hand, there are also strong arguments (and even stronger intuitions) which suggest that conscious states must be *distinct* from any material states.

I believe that in the end the materialist argument wins. Conscious states are material states. This is not to belittle the anti-materialist arguments and intuitions. They are deep and important. We will not grasp consciousness properly unless we understand how to answer them. Still, I think that careful analysis will show that they are flawed, and that the right solution is to embrace materialism.

p. 15 I shall begin by putting the materialist argument on the table. It is ↳ worth taking some care about this, for there are a number of different defences of materialism on offer in the contemporary literature, and not all of them are equally compelling. However, I think that there is one definitive argument for materialism. I shall call this 'the causal argument', and the burden of this first chapter will be to develop this argument and distinguish it from some less effective defences of materialism.

There is a further reason for laying out the argument for materialism carefully. Many contemporary philosophers harbour grave suspicions about materialism. Thus some philosophers contend that the whole idea of materialism is somehow empty, on the grounds that there is no proper way of characterizing the 'physical' realm. (Crane and Mellor 1990, Crane 1991, Segal 2000). And others suggest that contemporary materialism about the conscious mind rests on nothing but fashion or prejudice, unsupported by serious argument (Burge 1993, Clark 1996).

I intend to show that these attitudes are mistaken. The question of how to define 'physical' in the context of the mind-brain debate does raise a number of interesting points, but there is no great difficulty about pinning down a sense precise enough for the purposes at hand. It will prove easier do this, however, after we have rehearsed the argument for materialism. Accordingly, I shall not worry about the meaning of 'physical' at this stage, but simply begin by outlining the case for materialism. Once we have seen what is at issue, it will become clearer how materialists can best understand the meaning of 'physical', and I shall return to this issue at the end of the chapter.

There is one terminological point which I do need to address at this point, however. When I do fix a meaning for 'physical' at the end of the chapter, I shall read this term in a relatively strict sense, as standing roughly for the kinds of first-order properties studied by the physical sciences. Under the heading of 'materialism', on the other hand, I shall include not only the doctrine that conscious states are identical with physical states in this strict sense, but also the doctrine that they are identical with 'physically realized functional states', or with some other kind of physically realized but not strictly physical states (these possibilities will
p. 16 be explained further in section 1.6 ↳ below). It is true that the causal argument can be read as supporting the stricter identification with physical states, and indeed this is how I shall first present it in the next section. But, as we shall see, the causal argument can also be construed as supporting the less strict identification of conscious states with functional or other physically realized states. Since both the strict and the less strict identifications tie conscious states constitutively to the physical world, few of the arguments in this book will require me to decide between them. So it will be useful to have a term which covers both options, and I

have adopted ‘materialism’ for this purpose. Correspondingly, a ‘material’ state will mean either a physical state in the strict sense or some functional or other physically realized state.

In addition to suspicions about the meaning of ‘physical’, there is the further allegation mentioned above, that contemporary materialism is nothing but a modish fad. I take the causal argument to be outlined in this chapter to rebut this allegation. The causal argument may not be conclusive, but it certainly shows that the case for materialism goes beyond mere fashion or prejudice.

Some may think that the charge of modishness is supported by historical considerations. Widespread philosophical materialism is a relatively recent phenomenon, largely a creature of the late twentieth century. This recent provenance may seem to support the accusation that contemporary materialism owes its popularity more to fashion than to any serious argument. ‘If the case is so substantial’, anti-materialists can ask, ‘how come it took so long for philosophers to appreciate it?’ I take this to be a good historical question. But I think there is also a good historical answer: namely, that a key premiss in the argument for materialism rests on empirical evidence that only became clear-cut during the course of the twentieth century.

However, I shall not complicate the analysis of this chapter by overlaying it with historical commentary. The issues are complicated enough without the added burden of tracing historical strands. Accordingly, this chapter will focus on the structure of the argument for materialism, not its history. For those who are
p. 17 interested in the historical dimension, the Appendix at the end of this book discusses ↳ the history of the causal argument, and in particular the question of why it has become persuasive only recently.

1.2 The Causal Argument

Let me now outline what I take to be the canonical argument for materialism. Setting to one side all complications, which can be discussed later, it can be put as follows.

> Many effects that we attribute to conscious causes have full physical causes. But it would be absurd to suppose that these effects are caused twice over. So the conscious causes must be identical to some part of those physical causes.

To appreciate the force of this argument, consider some bodily behaviour which we would standardly attribute to conscious causes. For example, I walk to the fridge to get a beer, because I consciously feel thirsty. Now combine this example with the thought that, according to modern physical science, such bodily movements are fully caused by prior physical processes in brains and nerves. The obvious conclusion is that the conscious thirst must be identical with some part of those physical processes.

Let me now lay out the above argument more formally. This will help us to appreciate both its strengths and its weaknesses.

As a first premiss, take:

(1) Conscious mental occurrences have physical effects.

As I said, the most obvious examples are cases where our conscious feelings and other mental states cause our behaviour.

Now add in this premiss (‘the completeness of physics’ henceforth):

(2) All physical effects are fully caused by purely *physical* prior histories.[2]

p. 18 In particular, this covers the behavioural effects of conscious causes to which our attention is drawn by premiss 1. The thought behind premiss 2 is that such physical behaviour will always be fully caused by physical contractions in your muscles, in turn caused by electrical messages travelling down your nerves, themselves due to physical activity in your motor cortex, in turn caused by physical activity in your sensory cortex, and so on.

At first sight, premisses 1 and 2 seem to suggest that a certain range of physical effects (physical behaviour) will have two distinct causes: one involving a conscious state (your thirst, say), and the other consisting of purely physical states (neuronal firings, say).

Now, some events are indeed overdetermined in this way, like the death of a man who is simultaneously shot and struck by lightning. But this seems the wrong model for mental causation. After all, overdetermination implies that even if one cause had been absent, the result would still have occurred because of the other cause (the man would still have died even if he hadn't been shot, or, alternatively, even if he hadn't been struck by lightning). But it seems wrong to say that I would still have walked to the fridge even if I hadn't felt thirsty (because my neurons were firing), or, alternatively, that I would still have gone to the fridge even if my neurons hadn't been firing (because I felt thirsty). So let us add the further premiss:

(3) The physical effects of conscious causes aren't always overdetermined by distinct causes.

Materialism now follows. Premisses 1 and 2 tell us that certain effects have a conscious cause and a physical cause. Premiss 3 tells us that they don't have two distinct causes. The only possibility left is that the conscious occurrences mentioned in (1) must be identical with some part of the physical causes mentioned in (2). This respects both (1) and (2), yet avoids the implication of overdetermination, since (1) and (2) no longer imply *distinct* causes.

1.3 The Ontology of Causes

The causal argument focuses on the way in which conscious occurrences operate as *causes*. It says that
p. 19 conscious *causes* must ↳ be identical to physical *causes*. However, there are different philosophical theories of causation, and in particular about the kinds of things that can feature as causes. On one view, causes are facts, or instantiations of properties. Candidate causes on this view would be *my being in pain*, or *my having active nociceptive-specific neurons*. On an opposed view, causes are basic particulars, or *events*, abstracted from any conscious or physical properties they might have. The causal argument as stated above will generate different conclusions, depending on which view of causation you adopt. In particular, it will generate a stronger conclusion on the former view, that causes are facts, than on the latter view, which has causes as basic particulars. Still, this will be of no great moment, since a rephrasing of the argument will still allow us to generate the stronger conclusion, even on the assumption that causes are basic particulars.

Let me take this a bit more slowly. I myself favour the view that causes are facts (cf. Mellor 1995). A restricted variant of this view, which will perhaps be more familiar to some readers, is that causes are instantiations of properties by particulars, or 'Kim-events' (cf. Kim 1973). In what follows, I shall standardly use the term 'state' to refer to this kind of item—that is, to the possession of a property by some particular. Now, on the view that causes are facts (Kim-events, states) the causal argument given above implies that conscious *properties* (being thirsty, say) must be identical with physical *properties* (having a certain brain feature). For, in requiring that conscious causes be identical with physical causes, the argument will now require that conscious facts (Kim-events, states)—such as that *I am thirsty*, say—are identical to certain physical facts (Kim-events, states)—*I have a certain brain feature*, say—and these two facts (Kim-events, states) cannot be identical unless the properties they involve—being thirsty, having that brain feature—are themselves identical.

The alternative view of causation is that causes are basic particulars (cf. Davidson 1980). Then the causal argument, as phrased above, won't itself carry you to the identity of conscious and physical properties, since the identity of conscious Davidson-events with physical ones requires only the far weaker conclusion that the relevant conscious and physical properties are instantiated in the same particular, not that the properties themselves are identical.

p. 20 Still, as I said, we can rephrase the argument so as to regenerate the stronger conclusion. Let us take premiss 1′ to be the claim that all conscious events cause some physical events *in virtue of* their conscious properties; premiss 2′ says that all physical events are caused by prior physical events *in virtue of* the latter's physical properties; and premiss 3′ says that the physical effects of conscious causes aren't always caused twice over, *in virtue of* two different properties of the prior circumstances. In order to make these consistent, we then need once more to identify the conscious properties of the causes with their physical properties.

The causal argument as presented in the last section thus argues for the identification of conscious properties with physical properties. It is worth nothing at this stage, however, that this argument for property identity proceeds on an abstract, existential level, and is not concerned with any detailed identifications. It tells us that each conscious property must be identical with *some* physical property, but it doesn't tell us *which* specific physical property any given conscious property may be identical with.[3]

To establish any such specific property identity, more detailed empirical information is needed. It is not enough to know that conscious causes can always be identified with *some* part of the full physical histories behind their effects. To pin down specific property identities, we need more detailed evidence about correlations between specific conscious properties and the different parts of those physical histories. We need to know that pain, say, or thirst, or seeing an elephant, are found when such-and-such brain areas are active, but not when others are. In Chapter 7 I shall consider this kind of detailed research, and the kinds of results it can be expected to bring. But for the moment, I shall concentrate on the more abstract existential claim that every conscious property must be identical with *some*, as-yet-to-be-identified physical property.

p. 21 We can usefully think of this abstract claim and the detailed correlational research as complementing each other. The abstract claim doesn't by itself tell us which physical property a given conscious property should be paired up with. And the correlational research, while promising to establish specific pairings, can't by itself establish that the paired properties are *identical*, as opposed to regularly accompanying each other. The abstract claim is important, then, since it is needed to license the move from detailed empirical correlations to property identifications. It tells the empirical researchers that conscious properties aren't just *correlated* with the physical properties they are regularly found with, but must be identical with them.

1.4 Epiphenomenalism and Pre-Established Harmony

All this assumes, however, that the abstract claim does follow from the causal argument. Let us now examine this argument more closely.

As laid out above, the causal argument seems valid[4] So, to deny the conclusion, we need to deny one of the premisses. All of them can be denied without contradiction. Indeed, all of them have been denied by contemporary philosophers, as we shall see. At the same time, they are all highly plausible, and their denials have various unattractive consequences.

Let me start with premiss 1. This claims that, as a matter of empirical fact, particular conscious states have particular physical effects. This certainly seems plausible. Doesn't my conscious thirst cause me to walk to the fridge? Or, again, when I have a conscious headache, doesn't this cause me to ingest an aspirin?

Still, the possibility of denying this premiss is familiar enough, under the guise of 'epiphenomenalism' or 'pre-established harmony'.

p. 22 The first philosopher to embrace this option was Leibniz. Unlike ↳ most other philosophers prior to the twentieth century, Leibniz was committed to the causal completeness of physics (see Appendix). But he was not prepared to accept the identity of mind with brain. So he opted for a denial of our premiss 1, and concluded that mind and matter cannot really influence each other, and that the appearance of interaction must be due to *pre-established harmony*. By this Leibniz meant that God must have arranged things to make sure that mind and matter always keep in step. In reality, they do not interact, but are like two trains running on separate tracks. But God fixed their starting times and speeds so as to ensure they would always run smoothly alongside each other.

Some contemporary philosophers (for example, Jackson 1982) follow Leibniz in avoiding mind-brain identity by denying premiss 1. But they prefer a rather simpler way of keeping mind and matter in step. They allow causal influences 'upwards' from brain to mind, while denying any 'downwards' causation from mind to brain. This position is known as *epiphenomenalism*. It respects the causal completeness of physics, in that nothing non-physical causally influences the physical brain. But it avoids the theological complications of Leibniz's pre-established harmony, by allowing the brain itself to cause conscious effects.

Epiphenomenalism is not a particularly attractive position. For a start, it would require us to deny many apparently obvious truths, such as that my conscious thirst caused me to fetch a beer, or that my conscious headache caused me to swallow an aspirin. According to epiphenomenalism, my behaviour in both these cases is caused solely at the physical level. These physical causes may be accompanied by conscious thirst or a conscious headache, but these conscious states no more cause resulting behaviour than falling barometers cause rain.[5]

p. 23 That epiphenomenalism has these odd consequences is not in itself decisive. The theoretical truth can often overturn claims which were previously regarded as the merest common sense. Moreover, there is nothing incoherent about epiphenomenalism. As I shall have occasion to stress in what follows, there is nothing conceptually contradictory in the idea of conscious states which exert no causal powers themselves. Still, epiphenomenalism is surely an empirically implausible position, by comparison with the materialist view that conscious states are simply identical to brain states.

If epiphenomenalism were true, then the relation between mind and brain would be like nothing else in nature. After all, science recognizes no other examples of 'causal danglers', ontologically independent states with causes but no effects. So, given the choice between epiphenomenalism and materialism, standard principles of scientific theory choice would seem to favour materialism. If both views can accommodate the empirical data equally well, then ordinary scientific methodology will advise us to adopt the simple view that unifies mind and brain, rather than the ontologically more profligate story which has the conscious states dangling impotently from the brain states.

There remains the possibility that the anti-materialist arguments to be examined later will show that conscious mind and brain *cannot* be identical. If this is so, then one of the premisses of the causal argument must be false. And in that case premiss 1 seems as likely a candidate as any. Certainly most contemporary philosophers who are persuaded by the anti-materialist arguments have opted for epiphenomenalism and the denial of premiss 1, rather than for any other way out of the causal argument.

p. 24 But this does not invalidate the criticisms I have levelled against epiphenomenalism. My concern at the moment is not to prejudge the anti-materialist case, but merely to assess the causal argument. And the point remains that, in the absence of further considerations, it seems clearly preferable to identify mind with brain than to condemn conscious states to the status of causal danglers. It may be that further anti-

materialist considerations will yet require us to reconsider this verdict, but so far we have seen no reason to deny premiss 1, and good reason to uphold it.

Before leaving the issue of epiphenomenalism, it may be worth addressing some more local worries about premiss 1. Even if the blanket epiphenomenalist refusal to credit *any* conscious states with physical effects is methodologically unattractive, there may be some more specific reasons for doubting whether particular sorts of conscious states have the physical effects they are normally credited with. In particular, I am thinking here of conscious *decisions*, and doubts about their causal efficacy arising from the experimental results associated with Benjamin Libet, and of conscious states which are *representational*, and doubts about their causal efficacy arising from the possibility that they may have 'broad contents'. Let me deal with these in turn.

In a series of well-known experiments, Libet asked subjects to decide spontaneously to move their fingers, and simultaneously to note the precise moment of their decision, as measured by a large stop-watch on the wall. Libet also used scalp electrodes to detect the onset of motor cortical activity initiating the finger movement. Amazingly, he found that this neural activity started a full $\frac{1}{3}$ to ½ second *before* the subjects were aware of making any conscious decision (Libet 1993).

At first sight, this certainly suggests that such conscious decisions are epiphenomenal with respect to the actions we normally attribute to them: since the conscious decisions come later, it looks as if they must be effects of, rather than identical with, the brain processes that give rise to the action. But in fact this interpretation is not clear-cut. Libet himself points out that the conscious decisions still have the power to 'endorse' or 'cancel', so to speak, the processes initiated by the earlier cortical activity: no action will result
p. 25 if the action's ↳ execution is consciously countermanded. Given this, it seems that the conscious decision is part of the cause of the finger movement after all. The initial cortical activity does not determine the finger movement on its own, but only puts the motor cortex in a state of 'readiness', which leads to action in just those cases where the conscious decision is added. This then allows us to reason, as before, via the causal argument, that conscious decisions could not play a part in so influencing physical movements, were they not themselves physical.

In any case, even if conscious decisions did *not* contribute causally to the actions normally attributed to them, it would not follow that they had no physical effects of *any* kind. For instance, they will still presumably be causes of the sounds I make, or the marks I put on paper, when I later *report* my earlier conscious decisions. So they will still satisfy premiss 1, which requires only that conscious causes have *some* physical effects, and not that they have all the physical effects with which they are normally credited by common sense. So once more the causal argument will run.

The other worry concerned the possibility of conscious states with 'broad' representational contents. The possession of such 'broad contents' hinges on matters outside subjects' heads. For example, Hilary Putnam suggests that the representational state *thinking about water* hinges on what natural kind is actually water in your environment, and Tyler Burge argues that *thinking about arthritis* hinges on facts about other members of your community (Putnam 1975, Burge 1979, 1982).

Now the worry, in the present context, is that if any conscious states are representational in this broad way, then this will not sit happily with premiss 1's claims about causal efficacy. For how can states which hinge on matters outside your head exert a causal influence on your bodily movements? Surely your bodily movements are causally influenced solely by matters inside your skin, not by how matters are outside you.

The possibility of broadly representational conscious states raises any number of tricky issues, not all of which I can pursue here (though see section 7.7 below). However, they seem to me to pose no real threat to the causal argument for materialism. Let me content myself with two comments.

p. 26 First, I am open to the possibility that some, indeed all, conscious states may be essentially representational (cf. n. 1 above); moreover, it seems plausible that representation in general is a broad matter. Even so, it would seem odd to allow that conscious properties in particular, as opposed to representational properties in general, can depend on broad matters outside the skin. Could two people really be internally physically identical, yet nevertheless *feel* different, because things are different outside them? (Cf. Introduction, n. 2.) Given this, the natural strategy for those who seek to equate some (or all) conscious properties with representational properties is to shear off some species of narrow representation from the general run of broad representational properties, and to equate representational conscious properties with these narrow representational properties. And then, to return to the matter at hand, there will cease to be any reason to doubt that these conscious properties have physical effects such as bodily movements, however it may be with representational properties in general.

Second, even if you do wish to insist that some conscious properties are indeed broadly representational (a possibility to which I shall return in section 7.7), it will not follow that such broad conscious properties do not cause *any* physical effects. For they may have physical effects *outside* my body. For example, my consciously thirsting for water might affect *which liquid* I put into a glass, and my consciously worrying about arthritis might affect *where the doctor will poke me* when I complain of it. If this is right, then the causal argument will run as before, and imply that any such broad conscious properties must also be identical with physical properties, if their instantiations are to have such physical effects—though these physical properties will now presumably stretch outside bodies, as well as inside.

1.5 Accepting Overdetermination

There remain the two other premisses to the causal argument. It will be convenient to relegate the
p. 27 discussion of premiss 2, the completeness of physics, to the last section of this chapter and the ↳ Appendix. So let me now briefly consider premiss 3, the one ruling out overdetermination.

To reject this premiss is to accept that the physical effects of mental causes are always overdetermined by distinct causes. This is sometimes called the 'belt and braces' view (make doubly sure you get the effects you want), and is defended by D. H. Mellor (1995: 103–5).

At first sight, this position seems to have the odd consequence that you would still have gone to the fridge for a beer even if you hadn't been thirsty (because your cortical neurons would still have been firing), and that you would still have gone to the fridge even if your cortex hadn't been firing (because you would still have been thirsty). These counterfactual implications seem clearly mistaken.

However, defenders of the belt and braces view maintain that such implications can be avoided. They argue that the distinct mental and physical causes may themselves be strongly counterfactually dependent (that is, they hold that, if you hadn't been thirsty, your sensory neurons wouldn't have fired either, and vice versa).

Still, this then raises the question of *why* such causes should always be so counterfactually dependent, if they are ontologically distinct.[6] Why wouldn't my neurons have fired, even in the absence of my conscious thirst? Similarly, why shouldn't I still have been thirsty, even if my neurons hadn't fired? Now, it is not impossible to imagine mechanisms which would ensure such counterfactual dependence between distinct causes. Perhaps the conscious thirst occurs first, and then invariably causes the cortical activity, with both causes thus available to overdetermine the behaviour. Alternatively, the cortical activity could invariably cause the thirst. Or, again, the conscious decision and the cortical activity might be joint effects of some prior common physical cause. But such mechanisms, though conceptually coherent, seem highly

p. 28 implausible, especially given that they need to ↳ ensure that the conscious state and the brain state *always* accompany each other.

The relevant point is analogous to one made in the last section. We don't find any 'belt and braces' mechanisms elsewhere in nature—that is, mechanisms which ensure that certain classes of effects invariably have two distinct causes, each of which would suffice by itself. As with the epiphenomenalist model, a belt and braces model requiring such peculiar brain mechanisms would seem to be ruled out by general principles of scientific theory choice. If the simple picture of mental causation offered by materialism accommodates the empirical data as well as the complex mechanisms required by the belt and braces option, then normal methodological principles would seem to weigh heavily against the belt and braces view.

As with the corresponding argument for epiphenomenalism, this appeal to principles of scientific theory choice is defeasible. Perhaps in the end the anti-materialist arguments will force us to accept mind-brain distinctness. In that case, the belt and braces view might be worth another look. True, it is even more Heath-Robinsonish than epiphenomenalism. On the other hand, it does at least have the virtue of retaining the common-sense view that conscious states characteristically cause behaviour. In any case, my present purpose is not to decide this issue finally, but only to point out that, as things stand so far, we have good reason to uphold premiss 3, and none to deny it.

1.6 Functionalism and Epiphobia

Many contemporary philosophers will feel that the causal argument as elaborated so far is rather too strong. This argument has claimed that conscious properties are identical to *physical* properties. But the majority of contemporary materialists would probably prefer to identify mental properties in general, and conscious properties in particular, with physically realized *functional* properties, or properties which *supervene* on physical properties, or perhaps properties which are *disjunctions* of physical properties, rather than with strictly physical properties themselves.

p. 29 Let me start with functional properties. I shall come back to the other possibilities in a moment. A functional property is a higher-order property-of-having-some-property-which-satisfies-condition-R, where R specifies some requirement on an instantiation of a first-order property. In line with this, 'functionalism' in the philosophy of mind is the view that any given mental property should be identified with some property-of-having-a-first-order-property-which-bears-certain-causal-relationships-to perceptual inputs, behavioural outputs, and other mental states. For example, the property of being in pain might be identified, at first pass, with the property-of-having-some-property-which-arises-from-bodily-damage-and-gives-rise-to-a-desire-to-avoid-the-source-of-that-damage.

The advantage of this functionalist account of mental states is that it allows beings who have quite different intrinsic physical properties nevertheless to share mental properties. For example, it seems plausible that octopuses, whose neurology is physically quite different from human neurology, can nevertheless share the property of being in pain with humans. But, if this is so, the property of being in pain cannot be identical with any physical property, for no suitable physical property will be common to humans and octopuses. On the other hand, both humans and octopuses will share the higher-order property-of-having-some-property-which-arises-from-bodily-damage-and-gives-rise-to-a-desire-to-avoid-the-source-of-that-damage. The physical properties which play this role will be different in the two cases, but the higher-order property itself will be common to octopuses and humans.

Now, how does functionalism stand with respect to the causal argument? If we take the causal argument at face value, then they seem inconsistent. For, as we have seen, the causal argument promises to establish

that conscious properties are identical with strictly physical properties, which is just the claim that functionalism is designed to avoid.

This tension with the causal argument puts functionalism under some pressure. If functionalism is inconsistent with the causal argument, it must deny one of its premisses. And, on reflection, it could well be
p. 30 held to deny premiss 1, the one that says that conscious ↳ causes have physical effects. For, if conscious properties are not identical with physical properties, but rather with certain higher-order properties, then conscious causes will not be identical with the physical causes which premisses 2 and 3 tell us are the only causes of behavioural effects. So it would seem to follow that conscious states don't cause behavioural effects after all. This line of thought is sometimes said to generate 'epiphobia', a condition in which functionalists are overcome with anxiety about how their view differs from epiphenomenalism.

For this reason, and perhaps others, some philosophers have recently become uneasy about 'higher-order' properties. They object that it is profligate to posit substantial new properties for every way of *characterizing* objects as possessors of some (first-order) property which R (cf. Kim 1998: ch. 4).

I have some sympathy with this point of view. However, it is important to realize that, even if we reject higher-order properties on these grounds, the underlying dilemma highlighted by functionalism remains. For we will still need to decide whether conscious properties should be identified (a) with those strictly physical properties whose instantiations are paradigm physical causes, yet are not shared by humans and octopuses, or (b) with other first-order properties of a kind which can be shared by humans and octopuses, but are in danger of being outcompeted as serious causes.

Suppose, to illustrate the point, that we admit no properties except genuinely first-order properties. But suppose that we also continue to feel the pull of the thought that both humans and octopuses can be in pain. Given the physical differences between humans and octopuses, we might seek to respect this thought by construing *pain* as a *disjunctive* condition, requiring P_1 *or* P_2 *or* . . . where the various P_is are the different strictly physical properties which are causally active when different beings are in pain. But now epiphobia returns to trouble us once more. For my human arm movement is presumably caused by my human P_1 (my nociceptive-specific neurons firing, say). But P_1 itself isn't identical with the disjunction P_1 *or* P_2 *or* . . . — that is, with pain. So, if P_1 causes my movement, the disjunction presumably doesn't, and thus it seems to
p. 31 follow once more that the property of being in pain is inefficacious. The dilemma ↳ remains: if you want to have different creatures sharing pain, then you seem to end up rendering pains causally inefficacious.

A similar point can be made about views which replace higher-order functional properties, not by disjunctions of physical properties, but by properties which 'supervene' on physical properties. (For readers unfamiliar with this notion, it is explained in section 1.8.) This alternative will again leave us with the choice between identifying conscious properties with (a) physical properties themselves, or (b) with the properties which supervene on physical properties. And again we will face the dilemma that only (b) seems to allow physically different beings to share conscious properties, but only (a) seems to allow conscious properties to be causally efficacious.

In the next section I shall consider whether this dilemma can be resolved. However, it would be tiresome to have to address the issue separately for all the different ways in which conscious properties can be identified with properties which are not strictly physical—that is, for functional higher-order properties *or* disjunctions of physical properties *or* supervenient properties. So let me adopt the general term 'higher' property to cover all these alternatives. Correspondingly, when I speak of a 'higher' property being 'realized' by a physical property, I shall mean either that a functional higher-order property is instantiated because some physical property is, or that a disjunction of physical properties is instantiated because one of its disjuncts is, or that a supervenient property is instantiated because some physical property which determines it is.[7]

p. 32

1.7 A Possible Cure for Epiphobia

Perhaps there is a cure for epiphobia. We don't have to agree that the only respectable kind of *causation* involves strictly physical causes having physical effects. For it is arguable that there is a perfectly normal sense of 'cause' in which higher states cause the effects that their realizers cause. On this account, even if pain is a higher property, differently realized in octopuses and humans, my taking an aspirin can still be caused by the pain in my head, in virtue of being caused by whichever strictly physical state realizes that pain in me.

If we adopt this generous notion of causation, functionalism becomes consistent with premiss 1 of the causal argument after all. The fact that mental states are not identical with strictly physical states does not mean that they cannot cause the behaviour which is caused by those strictly physical states. In the generous sense of 'cause', they will do so as long as they are higher states which are realized by those strictly physical states.

Indeed, if we look at things in this way, we in effect have another version of the causal argument, one which reads 'cause' generously throughout, and which ends up with the conclusion that conscious properties, if not strictly physical properties, must at least be *physically* realized higher properties. The argument now runs:

(1*) Conscious causes have physical effects, at least in the generous sense.

(2) All physical effects are fully caused by purely physical prior histories.

(3) The physical effects of conscious causes aren't overdetermined by distinct causes.

And the conclusion is now that:

(4*) Conscious causes must at least be higher states which are realized by the *physical* causes of their physical effects.

For otherwise (3) would be violated, with the physical effects of conscious causes being caused twice over,
p. 33 first by their conscious ↳ causes as in (1′), and second by the distinct physical causes guaranteed by (2).

Note here how we do *in a sense* end up with two causes of the relevant behavioural effects. For we now have both (a) the higher state with which we are now identifying the conscious state and (b) the realizing physical state which directly causes the behavioural result. (Cf. Segal and Sober 1991.)

But the important point is that these two 'causes' are not now ontologically *distinct*, and so do not genuinely overdetermine any resulting behaviour. The higher cause is present only in virtue of the physical cause which realizes it. In the circumstances, the one would be absent if the other were. And because of this, we have no trouble with the counterfactuals which would be indicative of genuine overdetermination. It is *not* true that the behavioural result would still have been caused even if the physical realizer had been absent, for the higher state would then have been absent too;[8] and similarly, if the higher state had been absent in some particular case, there would again have been no alternative cause for the behavioural result, since the physical realizer would have had to be absent too.

Note that it is not essential to this rejigged version of the causal argument that we *start* with any assumption that conscious states are higher states. I shall be considering alternative arguments for materialism shortly, and in particular a form of argument that begins with a functionalist assumption of just this sort, taken to be derivable a priori from the structure of our concepts of conscious states. If you begin with this kind of a priori functionalism, a variant of the causal argument can still serve an important purpose: namely, that of establishing that higher mental states are *physically* realized, as opposed to being realized by some

p. 34 distinctive non-physical ↳ mind-stuff (cf. Lewis 1966). But this is not how I am thinking of the rejigged causal argument.

Rather, I intend it to establish *both* that conscious states must at least be higher states, if not strictly physical, *and* that they must be physically realized. That is, I am taking the identity of conscious properties with higher (or physical) properties to be the *conclusion* of my argument, not a premiss. The premisses are simply (1*), (2) and (3), which make no claims, a priori or otherwise, about the specific nature of conscious states, and the conclusion is that, if conscious properties are not strictly identical with physical properties, then they must at least be identical with higher-properties-which-are-physically-realized, otherwise we will be driven to deny that conscious states cause their effects in any sense or, alternatively, to accept that those effects are genuinely overdetermined by quite distinct causes.[9]

So far in this section I have shown how functionalism and other 'higher property' versions of materialism can respect the premisses of the causal argument, and indeed can use the rejigged version as an argument in their favour. However, I have not intended this as a defence of such views. This is because I am not sure whether they can really be cleared of the charge of epiphenomenalism.

The issue here hinges on whether we can seriously allow that higher states *cause* what their realizers cause. I am not sure what to say about this. Sometimes I think that this is not a serious notion of causation, and certainly not one which does justice to the way in which my thirst causes me to drink a beer. Surely, one feels, my thirst itself is efficacious in getting me to move, in just the same strict way as physical causes produce their effects, and not merely in the second-hand sense that it is realized by some other state which causes in this strict sense.

When I am in this mood, I am inclined to read the causal argument as employing a strict notion of causation
p. 35 throughout, and in particular in premiss 1's assertion that conscious states *cause* physical ↳ effects. This then drives me to the conclusion that conscious properties must be identical to strictly physical properties, and that any higher properties are merely epiphenomenal. The cost of this strictly physicalist position, of course, is that I will not share conscious properties with octopuses or other physically distinct beings. But perhaps this isn't as bad as it seems. After all, it doesn't mean that octopuses don't have any conscious properties at all. And I will still share *some* properties with them, albeit not the conscious properties that strictly cause our respective behaviours. (In particular, I will share some higher properties, which is perhaps why we can both count as in 'pain'.)

At other times I feel less fussy about causation. In particular, I sometimes worry that we will be left with precious few causes, if we are going to hold that higher states are pre-empted as causes whenever they have realizers in virtue of which they cause. For, if applied strictly, this principle threatens to block the causal efficacy of even such eminently respectable causal states as pressures and temperatures. After all, on any particular occasion the effects of temperatures and pressures will also be caused by specific molecular movements. These specific movements will *realize* the relevant pressures or temperatures, but won't be *identical* to them, since the pressures and temperatures can also be realized differently. So the pressures and temperatures won't count as causes, if they can't cause what their realizers cause.

This seems odd, and argues against dismissing higher states from the realm of serious causes, and in favour of a generous reading of premiss 1. On this reading, my thirst will still be a serious cause of my going to the fridge, even if it has a realizer in virtue of which it causes. And then the causal argument will simply yield the conclusion that it must be a physically realized higher state, not that it must be strictly physical itself.

As I said, I am not sure what to say about this issue. It is a complicated matter, and it is not clear how best to resolve it. Fortunately, nearly all the arguments in the rest of this book will be insensitive to this issue. We can identify conscious properties either with strictly physical properties or with physically realized higher
p. 36 properties. Whichever choice we make, we will still have an identity ↳ between conscious properties and

properties which are innocent of any of the obscurities which surround consciousness. This is the important point, and beyond that it will not matter too much whether conscious properties are identified with strictly physical or with physically realized higher properties.[10]

1.8 Intuition and Supervenience

Let me now distinguish the causal argument we have been examining from some other ways of defending materialism that can be found in the recent literature.

To start with, it is sometimes suggested that materialism about consciousness can be established by a priori intuition alone. This is a feeble thought, as will become clear shortly, but its deficiencies have sometimes been obscured by the fashion for thinking of materialism about the mental in terms of 'supervenience': that is, in terms of the doctrine that any two beings who share all physical properties must also share all mental properties.

I myself find the notion of supervenience more trouble than it's worth. The notion of supervenience has proved far less straightforward than it at first seemed, and has generated a huge amount of technical literature (mostly focusing on the 'must' in 'if . . . physically identical . . . *must* also be mentally identical'). I would argue that any benefits offered by the notion of supervenience are more easily gained simply by identifying mental properties directly with higher-order properties or disjunctions of physical properties. Accordingly, the notion of supervenience will not play a prominent part in the rest of this book.

p. 37 I mention it here only because supervenience formulations of ↳ materialism can create the spurious impression that materialism is a purely intuitive matter. After all, there is a sense in which a priori intuition does tell us that the conscious realm supervenes on the physical realm. Everybody has strong intuitions about the correlation between mind and brain. If I made a molecule-for-molecule physical copy of you using a Star Trek-style teletransporter, for example, wouldn't your physical twin automatically have all the same feelings that you have?

However, this intuition-based supervenience falls far short of anything worth calling materialism. To see why, note that the teletransporter thought-experiment is consistent with epiphenomenalism: perhaps the copy feels like the original simply because its brain states causally generate extra conscious states, in just the same way as the original's brain states do. Here the conscious states would be distinct from the brain states, but would regularly accompany them, in virtue of laws by which brain states cause conscious states. A merely epiphenomenalist mind-brain correlation like this clearly doesn't amount to materialism.

In the technical terminology into which we are forced by the apparatus of supervenience, the point is that the teletransporter thought-experiment shows only that physical identity guarantees conscious identity across *natural* possibilities, possibilities which share all our natural laws, including any brain-mind epiphenomenal laws. However, to establish a supervenience amounting to genuine materialism, we would need to show that physical duplicates couldn't *possibly* be mentally different, whatever the laws of nature, not just that they aren't different in worlds which do share our laws. We need to establish supervenience of the mental across all metaphysically possible worlds. Only this promises to ensure that the mental is ontologically inseparable from the physical, and not just correlated with it.

If you find this obscure, the point can be put more directly in terms of ontological relations between mental and physical properties. Mere supervenience across naturally possible worlds doesn't amount to materialism, because it doesn't rule out the epiphenomenalist possibility that conscious properties are
p. 38 ontologically quite distinct from physical properties, albeit constantly correlated with them by ↳ epiphenomenal laws in this actual world and those nearby worlds that share our natural laws.

Supervenience across all possible worlds, on the other hand, does arguably suffice for materialism, precisely because an ontological dependence of mental on physical properties seems the only thing that will enable physical identity to *necessitate* mental identity, whatever laws may obtain.

Now that we see which version of supervenience is required to ensure genuine materialism, it should be clear that intuition alone will fail to deliver the materialist goods. It is not at all intuitively obvious that physical duplicates must *necessarily* be conscious duplicates, that a physical doppelganger couldn't *possibly* have different experiences. Even a dyed-in-the-wool materialist, like myself, feels the pull of the intuition that there could be a 'zombie', say, who is physically just like me but has no feelings—in a possible world, so to speak, where any epiphenomenal laws relating brain states to conscious states have broken down.

It may in fact be true that zombies are impossible, and indeed this is something for which I shall argue at length in due course. My present point is only that a priori intuition alone cannot establish their impossibility. If anything, it suggests just the opposite.

1.9 An Argument from a Priori Causal Roles

Let me now consider one further form of argument for materialism. This shares some of the structure of the causal argument. But in place of premiss 1 or 1^*, which simply states that, as a matter of fact, conscious causes have physical effects, this argument appeals instead to a putative a priori analysis of our *concepts* of conscious states.

According to this line of thought, our concepts of conscious states, like *pain*, or *thirst*, are each associated a priori with the specification of some causal role linking that state to physical causes and effects (cf. Lewis 1966). So, as above, our concept of pain would be linked a priori with bodily damage as cause and a desire to avoid the source of the pain as effect. Again, our concept of thirst would be linked to lack of water as cause and a desire to drink as effect.

p. 39 This kind of a priori analysis can then be plugged into the rest of ↳ the causal argument, so to speak, to deliver the materialist conclusion. The a priori analysis tells us that conscious states have a causal role, and hence have physical effects. The completeness of physics tells us that these physical effects must have full physical histories. The denial of overdetermination tells us that these physical effects aren't caused twice over. Thus, once more we reach the conclusion that conscious states cannot be ontologically distinct from the physical causes of their physical effects.[11]

There may seem no great distance between the causal argument discussed earlier and this argument appealing to a priori analyses of our concepts of conscious states. However, it is crucially important that the causal argument discussed earlier rests on no such a priori assumptions. While that causal argument assumed that conscious causes have physical effects, it offered this as a straightforward empirical truth, not as a conceptual matter.

In line with this, note how I have been happy to allow the *conceptual* possibility that conscious states may lack effects altogether. This point arose earlier in my discussion of epiphenomenalism. My reason for dismissing epiphenomenalism was not that its denial of mental efficacy violated any conceptual truths, but simply that it amounted to an empirically far less plausible story than the simple identities postulated by materialism.

I shall have a lot more to say about our concepts of conscious states in what follows. Without wanting to pre-
p. 40 empt that analysis, let me simply say at this stage that it will amply confirm that there are no ↳ a priori associations between concepts of conscious states and specifications of causal roles.

On this conceptual issue, I am thus in agreement with a number of recent writers who have argued that the a priori style of argument for materialism doesn't work (cf. Levine 1983, Chalmers 1996). They object to the initial a priori claim about concepts of conscious states. Our concepts of conscious states are not a priori related to any specifications of causal roles, they protest. So there is no conceptual route, they conclude, from the fact that any causal roles must be filled by physical states to the conclusion that conscious states are material.

I accept this criticism of the a priori style of argument for materialism. To repeat, I agree that our concepts of conscious states are not associated a priori with causal roles. But this isn't as bad for materialism as Levine and Chalmers suggest. If the a priori argument were the *only* argument for materialism about consciousness, then materialism would indeed be in trouble. However, it is not the only argument. There is also the original causal argument as I have presented it, which does not depend on any particular assumption about our concepts of conscious states.

1.10 What Is 'Physics'?

Let me now address a terminological issue flagged earlier, an issue that may have been worrying readers for some time. How exactly is 'physics' to be understood in this context of the causal argument? An awkward dilemma may seem to face anyone trying to defend the crucial second premiss, the completeness of physics. If we take 'physics' to mean the subject-matter currently studied in departments of physics, discussed in physics journals, and so on, then it seems pretty obvious that physics is not complete. The track record of past attempts to list *all* the fundamental forces and particles responsible for physical effects is not good, and it seems highly likely that future physics will identify new categories of physical cause. On the other hand, if
p. 41 we mean by 'physics' the subject-matter of such ↳ future scientific theories, then we seem to be in no position to assess its completeness, since we don't yet know what it is.

This difficulty is more apparent than real. If you want to use the causal argument, it isn't crucial that you know exactly what a complete physics would include. Much more important is to know what it won't include. (Cf. Papineau and Spurrett 1999.)

Suppose, to illustrate the point, that we have a well-defined notion of the *mental* realm, identified via some distinctive way of picking out properties as mental. (Thus we might identify this realm as involving intentionality, say, or intelligence, or indeed as involving consciousness—the precise characterization won't matter for the point I am about to make.) Then one way of understanding 'physical' would simply be as 'non-mentally identifiable'—that is, as standing for properties which can be identified independently of this specifically mental conceptual apparatus. And then, provided we can be confident that the 'physical' in this sense is complete—that is, that every non-mentally identifiable effect is fully determined by *non-mentally identifiable* antecedents—then we can conclude that all mental states must be identical with (or realized by) something non-mentally identifiable (otherwise mental states couldn't have non-mentally identifiable effects).

This understanding of 'physical' as 'non-mentally identifiable' is of course a lot weaker than any normal pre-theoretical understanding, but note that it still generates a conclusion of great philosophical interest: namely, that all mental states, and in particular all conscious states, must be identical with non-mentally identifiable states. We may not know enough about physics to know exactly what a complete 'physics' might include. But as long as we are confident that, whatever it includes, it will have no ineliminable need for any distinctively mental categorizations, we can be confident that mental properties must be identical with (or realized by) certain non-mentally identifiable properties.

In fact, I shall understand 'physical' in a somewhat tighter sense in what follows, as 'identifiable non-mentally-*and*-non-biologically', or 'inanimate' for short, rather than simply as 'non-mentally identifiable'. This is because it is this realm, the 'inanimate', that is most naturally argued to be complete. When I
p. 42 examine the detailed ↳ scientific reasons for believing in the completeness of physics, in the Appendix, it will turn out that the realm which science has in fact shown to be causally sufficient unto itself is the inanimate. What science has actually shown is that any inanimate effect (that is, any effect specifiable in terms of mass, or charge, or chemical structure, or . . . in any non-biological and non-mental way) will have an inanimate cause. So it is this thesis that I propose to plug into the causal argument. Conscious causes have inanimate effects. Inanimate effects always have full inanimate causes. So conscious properties must be identical with (or realized by) inanimate properties.[12]

It might not be immediately obvious why I am being so careful here. Why not simply read 'physical' as non-mentally identifiable, as I suggested initially? If the Appendix succeeds is showing that the inanimate is complete, then won't it *a fortiori* show that the non-mentally identifiable is complete? After all, if something is inanimate, then it is certainly non-mentally identifiable. So, if the inanimate is complete, and there are inanimate causes for all inanimate effects, then those causes will be non-mentally identifiable too. And this would thus seem to ensure the completeness of the non-mentally identifiable.

No. This is too quick. To see why, take an *effect* which is *not* inanimate yet *is* non-mentally identifiable. An *arm* moving would be a good example. I take it that the notion of an arm movement is not a *mental* notion. But the notion of *an arm* is certainly a biological notion. So arm movements are not inanimate, even though they are non-mentally identifiable.

p. 43 Now, the completeness of the inanimate tells us that all inanimate ↳ effects have inanimate causes. But, since *arm* movements aren't inanimate, it doesn't follow that they have inanimate causes, nor, therefore, that they must have non-mentally identifiable causes. Maybe, for all the completeness of the inanimate guarantees, arm movements are always caused by mental states alone, like desires or intentions, without any assistance from further causes at the inanimate level. This thus shows that the completeness of the inanimate doesn't guarantee the completeness of the non-mentally identifiable.[13]

This last point suggests a possible way of resisting the causal argument. Anti-materialists could allow that there is a familiar everyday sense in which conscious states have 'physical' effects, and another good sense in which the 'physical' realm is complete, and yet object that the causal argument fails to go through because the two senses are distinct. (Cf. Sturgeon 1998.[14]) Thus they could allow that conscious causes always have non-mentally identifiable bodily effects like arms moving, and also allow that the seriously inanimate realm of mass and motion is complete, but urge that since arm movements aren't themselves inanimate, it doesn't follow that they must have inanimate causes, nor therefore that their conscious antecedents must be identical with anything inanimately identifiable.

This is a serious enough issue, but it is scarcely conclusive against the materialist side. Materialists need
p. 44 only make sure that their senses of 'physical' line up properly. The version of completeness I ↳ take to be defensible, as I said, is the completeness of the inanimate. So all I need to make the causal argument go through is a version of premiss 1 which will ensure that conscious causes do have inanimate effects, in addition to their effects on animate body parts.

One way of arguing for this premiss would be to start with the point that the anti-materialist concedes, namely, that conscious states cause animate effects, like arms moving, and then argue that arm movements should themselves be identified with inanimate occurrences, thus giving the conscious causes inanimate effects, as desired. An obvious strategy here would be to note that arm movements themselves have inanimate effects (such as stones flying through the air, say), and then apply the causal argument once

more, to conclude that these arm movements must be identified with the inanimate causes of those inanimate effects, if we are to avoid overdetermination.

But this is a somewhat long way round. If the animate bodily effects of conscious causes have inanimate effects, then we can infer directly that the conscious causes must themselves have those inanimate effects, by transitivity, whether or not the animate bodily movements are identified with inanimate occurrences. (Cf. Witmer 2000.) Thus, if some conscious desire causes my arm to move, and this movement in turn has such inanimate effects as a stone flying through the air, a window shattering, and so forth, then my conscious desire itself will cause these inanimate effects. I take it to be uncontroversial that conscious states standardly[15] have such inanimate effects, and will assume this henceforth.

1.11 The Completeness of Physics

Let me conclude this chapter with a few remarks about the causal argument's second premiss, the
p. 45 completeness of physics. It is one ↳ thing to fix a sense of 'physics' which renders this a substantial claim which might be true or false. It is another to show that it is in fact true.

Some readers might feel that this is not a problematic issue. Once we have fixed a definite meaning for 'physical', as equivalent to 'inanimate', say, then is it not just a matter of common sense that all physical effects will have physical causes? In particular, if we take the physical effects in this sense that we normally attribute to conscious causes, then is it not obvious that these effects can always in principle be fully accounted for in terms of uncontroversially physical histories, involving the movement of matter (in arms), molecular processes (in muscles), the action of neurotransmitters (in brains) . . . and so on?

This is certainly how I thought of the issue when I first started working on the causal argument. I realized that this argument involved a number of disputable moves, and was therefore ready for it to be queried on various different grounds. But the one assumption that I did expect to be uncontroversial was the completeness of physics. To my surprise, I discovered that a number of my philosophical colleagues didn't agree. They didn't see why some physical occurrences, in our brains perhaps, shouldn't have irreducibly conscious causes.

My first reaction to this suggestion was that it betrayed an insufficient understanding of modern physics. Surely, I felt, the completeness premiss is simply part of standard physical theory. However, when my objectors pressed me, not unreasonably, to show them where the completeness of physics is written down in the physics textbooks, I found myself in some embarrassment. Once I was forced to defend it, I realized that the completeness of physics is by no means self-evident. Indeed, further research has led me to realize that, far from being self-evident, it is an issue on which the post-Galilean scientific tradition has changed its mind several times. The completeness of physics may seem the merest part of common sense to many of us today, but as recently as 150 years ago most people, including most orthodox scientists, would have thought the idea absurd, taking it to be obvious that there must be some *sui generis* conscious states in the causal history of human behaviour.

p. 46 So the completeness of physics is a doctrine with a history, and a very interesting history at that. In the Appendix I detail this history. My main purpose in doing this is to show that there is good empirical evidence for the completeness of physics. But the historical story also shows that this evidence is relatively recent, and that prior to the twentieth century the empirical case for the completeness of physics was by no means persuasive.

At the beginning of this chapter I raised the question of why philosophical materialism has become popular only in the last fifty years or so. As I pointed out, this historical circumstance lends weight to the suggestion

that contemporary materialism is a creature of fashion rather than serious philosophical argument. I take the story I tell in the Appendix to rebut this suggestion. There is indeed a good case for materialism. But it has not always been available to philosophers. This is because its crucial premiss, the completeness of physics, rests on empirical evidence which has emerged only relatively recently.

Notes

1 Some philosophers assume that 'phenomenal' is meant to contrast with *intentional*, and on this basis hold that much recent discussion of consciousness, especially that surrounding David Chalmers's 'hard problem' (1996), is invalidated by an implicit supposition that subjectivity is independent of intentionality (cf. Eilan 1998). It is perhaps worth emphasizing that I don't intend 'phenomenal' to imply 'non-intentional'. I simply mean it as a non-committal term for subjective 'what-it's-likeness'. Nothing yet rules out the possibility that all, or only, intentional states involve phenomenal consciousness. Moreover, since Chalmers also understands 'phenomenal' in this way, his 'hard problem' of phenomenal consciousness will still arise even if phenomenality is not independent of intentionality.

2 What about quantum indeterminacy? A stricter version of (2) would say that the *chances* of physical effects are always fully fixed by their prior physical histories, and would reformulate the rest of the argument accordingly (with (1) then as 'Conscious mental occurrences affect the *chances* of physical effects', and so on). I shall skip this complication in most of what follows.

3 Similarly, on the alternative construal of the causal argument to be developed in section 1.7, we will have an abstract argument for the identity of conscious properties with material properties (even if not with strictly physical properties), but again this argument on its own will not tell us which material properties any specific conscious property should be identified with.

4 However Sturgeon (1998) argues that the argument trades on an equivocation between the everyday sense of 'physical' (in premiss 1) and a quantum-theoretical sense (in premiss 2). I shall comment on Sturgeon's claim in section 1.10 below.

5 Chalmers (1996: esp. 134–6), following Russell (1927) and Lockwood (1989), argues that there is a way for dualism to avoid this epiphenomenalist inefficacy while respecting the completeness of physics. This is to identify phenomenal properties with the *intrinsic* properties of the physical realm. Chalmers's idea is that physical science picks out properties like mass and charge only extrinsically, via their relations to observable features of the world. So maybe phenomenal properties can be identified with the intrinsic nature of such properties, suggests Chalmers, and thereby have their causal efficacy restored. This seems an entirely sensible view to me. But, *pace* Chalmers, I would say that it is simply a version of materialism. My reaction is that the intrinsic features of the physical world with which Chalmers wants to identify phenomenal properties are themselves simply basic physical properties. Thus I am happy to agree with Chalmers that scientific theory picks out these intrinsic physical properties only via descriptions which refer to observable features of the world. Moreover, I agree that conscious properties should be identified with arrangements of such intrinsic physical properties, and thus that it is like something to have these arrangements of intrinsic properties. Indeed, I find it hard to see what a sensible materialism could amount to, except this combination of views. So, from my point of view, Chalmers's suggested position is simply the optimal formulation of materialism.

6 Note that this is only a problem if the causes are genuinely ontologically distinct, and not if they are merely related as role state and physical realizer. As we shall see in the next section, the existence of 'two' causes in this latter sense does not threaten overdetermination, precisely because of their ontological interdependence. So I have no objection to versions of the belt and braces view which intend only parallel causes in this weak sense. Cf. Segal and Sober 1991.

7 It might seem that for completeness I should also consider the possibility that conscious properties can be identified with macro-properties that are *composed of* physical micro-properties. However, the mereological notion of a macro-whole being composed of micro-parts seems to me orthogonal to the notion of one kind of property being *realized* by another kind of property. I would say that composition can occur *within* the strictly physical, and accordingly that a macro-property composed of strictly physical micro-properties is itself a strictly physical property. So identifying conscious properties with properties composed of strictly physical properties is itself to identify them with strictly physical properties, not with some species of 'higher' property realized by physical properties. There are interesting questions about the compositional relation between micro-parts and macro-wholes—in particular, do the parts have some kind of causal primacy over wholes?—but they are independent of the arguments in this book. (For convincing arguments against the causal primacy of micro-parts, see Hüttemann, forthcoming.)

8 Mightn't this be false? Since different physical properties can realize a given higher state, isn't possible that a different realizer would have been present if the actual realizer had been absent? There are some delicate issues here (cf. Yablo

1992). But for present purposes the significant point is that a given creature certainly wouldn't have had a given higher property if it hadn't had the kind of physical property that features as a realizer when it itself has that higher property. For example, I certainly wouldn't have been in pain if I didn't have the kind of physical property that realizes pain in me (since after all, there isn't any question of my having the realizer that occurs in octopuses or other creatures).

9 I take this construal of the causal argument to rebut Tim Crane's (1995) complaint that the causal argument must assume a form of mental causation which will then be denied by any functionalist-style higher-property version of materialism.

10 In Chapter 7 I shall return to the choice between functionalist-style higher-property views and strict physicalism, and consider whether empirical research into conscious properties can help resolve the issue. By that stage I shall also have developed an extensive account of the structure of our *concepts* of conscious properties. But I shall argue that none of this delivers a resolution, and my eventual conclusion will be that our concepts of conscious properties are vague, in that it is indeterminate whether they refer to higher properties or to the physical properties which realize them in humans.

11 Interestingly, this form of argument can be used to deliver either a functionalist version of materialism—conscious states are identical with physically realized higher-order states—or a more strictly physicalist conclusion—conscious states are identical with the physical states themselves. We will get the former conclusion if we take it to be given a priori that conscious properties are higher-order properties—that is, that they are identical to properties like the-property-of-having-some-property-which-plays-such-and-such-a-causal-role. The overall argument will then lead us to the conclusion that these higher-order properties must be realized *physically*, as opposed to being realized by some special mind-stuff. But we can also take the causal roles associated a priori with concepts of conscious states to fix reference to whichever first-order properties actually play those roles, rather than to the roles themselves. The a priori argument then delivers the conclusion that conscious properties are identical to strictly physical properties. This is in fact how the argument is run by David Lewis.

12 Note how 'animate' and 'physical $=_{df}$ inanimate' are not being used exclusively here. A property is 'animate' if it is identifiable pre-theoretically in mental or biological terms, as involving, say, intelligence, or consciousness, or respiration, or digestion. A property is 'inanimate' if it can be identified in some other way: paradigm 'inanimate' properties would thus be size, shape, mass, charge, and combinations thereof. This allows that certain properties can be both 'animate' and 'physical $=_{df}$ inanimate'. Such properties will be those that can be identified both ways: first, in 'animate' terms, and second, as (complexes of) 'inanimate' features which happen to be instantiated in minded or biological systems. This is of course the right way to set things up, given that we don't want definitions to rule out the materialist possibility that 'animate' properties are identical to 'inanimate' properties.

13 The point is that switching from inanimate to non-mentally identifiable not only gives us fewer potential causes to falsify completeness, but also more effects, since the relevant completeness thesis now covers extra animate yet non-mentally identifiable effects, like raisings of arms. So we cannot infer from the fact that there are no *sui generis* mental causes for any inanimate effects that there are no *sui generis* mental causes for effects like the raising of arms. I would like to thank Finn Spicer for helping me to see this clearly.

14 In fact, Sturgeon's charge is not that conscious causes fail to have inanimate effects, but rather that they fail to have quantum-mechanical effects. This is in line with his assumption that the materialists' causal argument will appeal to the causal completeness of quantum mechanics, rather than the causal completeness of the inanimate. However, as the Appendix will make clear, this isn't my completeness thesis. I don't think of quantum mechanics *per se* as asserting completeness, since the basic assumptions of quantum mechanics leave it open what forces (Hamiltonians) there are. Rather, my crucial completeness claim is that all inanimate accelerations are due to inanimate forces.

15 Mightn't *some* conscious occurrences, like wishful thinking, lack any inanimate effects? Well, I am happy to agree that the causal argument does not engage directly with such inefficacious states, and that this thus creates space for the possible view that specifically those conscious states are immaterial, while those with physical effects are not. But this does not seem a serious position to me. If anybody really wants to pursue it, they can send me an e-mail, and I'll think of some arguments.