sense of how they are caused and the laws governing them, or in fact why they exist at all. And that is why we can't.

It is not sufficiently appreciated that Physicalism is an extremely optimistic view of our powers. If it is true, we have, in very broad outline admittedly, a grasp of our place in the scheme of things. Certain matters of sheer complexity defeat us—there are an awful lot of neurons—but in principle we have it all. But consider the antecedent probability that everything in the Universe be of a kind that is relevant in some way or other to the survival of *Homo sapiens*. It is very low surely. But then one must admit that it is very likely that there is a part of the whole scheme of things, maybe a big part, which no amount of evolution will ever bring us near to knowledge about or understanding of. For the simple reason that such knowledge and understanding is irrelevant to survival.

Physicalists typically emphasize that we are a part of nature on their view, which is fair enough. But if we are a part of nature, we are as nature has left us after however many years of evolution and each step in that evolutionary progression has been a matter of chance constrained just by the need to preserve or increase survival value. The wonder is that we understand as much as we do, and there is no wonder that there should be matters which fall quite outside our comprehension. Perhaps exactly how epiphenomenal qualia fit into the scheme of things is one such.

1.3 The Case for Materialism

DAVID PAPINEAU

David Papineau teaches at King's College London and at the Graduate Center of the City University of New York. He has written widely in the philosophy of mind, metaphysics, and the philosophy of language.

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.

For the moment, 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. 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 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

David Papineau, *Thinking about Consciousness* (Oxford University Press 2002), pp. 14–18, 21–23, 27–28, 40–41, 45. By permission of Oxford University Press.

or subjective consciousness. 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.

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.

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.

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. And others suggest that contemporary materialism about the conscious mind rests on nothing but fashion or prejudice, unsupported by serious argument.

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 to 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.

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.

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.

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.

EPIPHENOMENALISM AND PRE-ESTABLISHED HARMONY

Let us now examine the causal argument more closely.

As laid out above, the causal argument seems valid. 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'.

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. 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 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.

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.

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.

ACCEPTING OVERDETERMINATION

There remain the two other premisses to the causal argument. It will be convenient to relegate the discussion of premiss 2, the completeness of physics, to the last section of this chapter. 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).

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 counter-factually 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. 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 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.

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 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.

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. 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.

THE COMPLETENESS OF PHYSICS

Let me conclude this chapter with a few remarks about the causal argument's second premiss, the

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.

So the completeness of physics is a doctrine with a history, and a very interesting history at that. 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.

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.

1.4 Functionalism and Eliminative Materialism

PAUL CHURCHLAND

Paul Churchland has written widely in the philosophy of psychology and related fields. He teaches at the University of California, San Diego.

From Churchland, Paul M., *Matter and Consciousness*, 2808 word excerpt from pages 36–48, © 1984 Massachusetts Institute of Technology, by permission of The MIT Press.