

## Confusions about Consciousness

*David Papineau*

### *Introduction*

Consciousness has suddenly become an extremely fashionable topic in certain scientific circles. Many thinkers are now touting consciousness as the last unconquered region of science, and theorists from many different disciplines are racing to find a 'theory of consciousness' which will unlock this final secret of nature. I am suspicious about all this enthusiasm. I think that much of the brouhaha is generated by philosophical confusion. In the end, I fear, there is no special secret of consciousness, and no special key needed to unlock it.

### *Dualism and Materialism*

Much of the confusion about consciousness is generated by lack of clarity on the issue of dualism. The majority of scientists who are caught up in the current excitement about consciousness studies would probably deny that they are dualists, if the question were put to them explicitly. But at the same time I think that many of them are closet dualists. They strive to resist the temptations of dualist thinking, but as soon as their guard drops they slip back into the old dualist ways. The very language in which they normally pose the problem of consciousness gives the game away. 'How can brain states "give rise" to conscious feelings?' 'How are conscious states "generated" by neural activity?' The way these questions are phrased makes it clear that consciousness is being viewed as something extra to the material brain, even if the official doctrine is to deny this.

To help make the point clear, let me move away from the mental realm for a moment, and consider two contrasting analogies from purely physical science, the theory of electromagnetism, and the theory of heat. These two theories work rather differently. Think of how they relate heat and electromagnetism to other physical processes. The theory of the electromagnetic field is a theory of an *extra* physical entity, of something additional to other physical goings-on, such as the movement of charged particles. The charged particles are one thing, and the field they produce something further. But the theory of heat does not explain heat in a similar way. Heat is not something extra to the kinetic energy of moving particles. Rather, talk of the heat in a body is just another way of referring to the kinetic energy of the particles in it. There aren't two entities here, the moving particles *and* the heat. It's not as if a 'heat field' arises when the particles move. Heat is nothing but the movement of the particles, described in other terms.

Now, which of these is the better model for the relation between conscious feelings and brain activity? That is, should we expect a successful 'theory of consciousness' to show us how certain brain activities *generate* certain extra entities, the conscious feelings, on the model of the electromagnetic field? Or should we rather expect such a theory to show us how conscious feelings *are nothing but* certain brain activities, described in other terms, on the model of heat. When Francis Crick, for example, says that consciousness is associated with 40-Hertz neuronal oscillations in the visual cortex, or indeed when any scientist equates consciousness with any feature of brain

activity, are we to understand them as saying that some extra conscious field is generated by the brain activity, or rather that consciousness is nothing but that brain activity, described in other terms?

We can call a theory of the former kind a dualist theory, and a theory of the latter kind a materialist theory. I suspect that much work in 'consciousness studies' simply hasn't decided whether it is aiming at a dualist theory or a materialist theory. The indecision matters because it can lend such work an air of spurious excitement. This is because a dualist theory of consciousness, while it would certainly be exciting, is a highly implausible prospect. A materialist theory, by contrast, while it is plausible enough, is not going to yield any exciting secrets. So, by fudging the issue between these two kinds of hypothesis, theorists of consciousness can have their cake and eat it. They can present their work as sharing the excitement of a dualist breakthrough, yet at the same time denying that its claims are any more surprising than a materialist hypothesis. If we are seriously to assess their theories, however, we need to be told whether they are intended in the dualist or materialist mode.

### *Consciousness and Life*

To further clarify this issue, let me turn to another analogy, this time with the 'theory of life'. About a hundred and fifty years ago, scientists used to be excited about life in roughly the way that they are now excited about consciousness. While they were of course clear enough about which living systems are alive and which not, they were much perturbed by further questions. Why are those systems alive? What mysterious power animates them? And why is this power present in certain cases, such as trees and oysters, and not in others, like volcanoes and clouds?

These questions have now disappeared from active debate. Biology textbooks sometimes begin with a few perfunctory paragraphs about the distinguishing characteristics of their subject matter. But the nature of life is no longer a topic of serious theoretical controversy. Everybody now agrees that the difference between living and non-living systems is simply having a certain kind of physical organisation (roughly, we would now say, the kind of physical organisation which fosters survival and reproduction).

The reason for the nineteenth-century debate, and its subsequent disappearance, is that scientists used to be dualists about life, and aren't any longer. That is, they used to think that living systems are animated by the presence of a special substance, a vital spirit, or *élan vital*, which was postulated to account for those features of living systems, such as generation and development, which were thought to be beyond physical explanation. And of course, when they did believe in these vital forces, they then faced any number of exciting questions, such as why they arise in certain circumstances and not others, and what laws govern their operation.

But nobody is a dualist about life any longer. Nobody believes in vital spirits nowadays. A century and more of physiological research have persuaded scientists that the characteristic features of living systems can all in principle be accounted for in terms of normal physical forces, without bringing in any extra forces operating only in living bodies. With this realisation all the excitement about the nature of life has dissolved. To be alive is just to be a physical system of a certain general physical

kind. There isn't any extra property present in living systems, over and above their physical features, which distinguishes them from non-living systems. So there are no pressing questions about the mysterious nature of this extra property.

I think that this story about life carries a direct moral for the study of consciousness. If you think that there are special mental forces, over and above the familiar physical forces, then you will think that there are exciting questions that must be answered, such as why these forces arise in certain circumstances and not others, and what laws govern their operation. On the other hand, if you think that the cognitive workings of intelligent beings depend on nothing but the operation of normal physical forces, without any extra forces operating only in brains, then you will see things differently. You may begin your textbooks with a few remarks about the distinguishing characteristics of conscious systems, but once this essentially classificatory question is out of the way, you won't want to spend any more time agonising about the nature of consciousness.

As my remarks so far will no doubt have intimated, I prefer the latter, materialist view of consciousness to the former, dualist story. And the reason is the same as in the case of life. All the physiological evidence indicates that no special mental forces are needed to account for the operation of intelligent organisms. Of course the evidence isn't conclusive, and doesn't absolutely prove that there are no such forces. But it weighs strongly against them. If you are unpersuaded, then ask yourself this question. Are any parts of matter in your brain ever caused to accelerate by mental causes, in the absence of any other forces? That is, do we need to include purely mental causes alongside gravity, the electroweak force, and so on, in the category of fundamental forces? As I said, there is no conclusive disproof of this thought, and it has its defenders, like Sir John Eccles (1989). But I take it that it would run counter to a large body of empirical evidence. (physicists would certainly be very interested if such a force could be shown to exist.)

### *Epiphenomenalism*

No doubt some of you will be feeling uneasy about the analogy with life. Don't we have immediate access to the nature of conscious activity, via our introspective knowledge of our own minds, in a way that we don't have access to the nature of life? And doesn't this show us directly that conscious goings-on are distinct from any physical goings-on? When we are aware of a pain, say, or of seeing something red, don't we know directly that there is something going on in us which is quite different from any neuronal activity in our brains? I agree that we all have strong intuitions to this effect. But they need to be handled with care. My own view is that they are illusory, and I will come to this in the next section. Still, some thinkers take these intuitions at face value, as showing that conscious feelings really are distinct from brain activity. However, if you take this line, then you face the argument from two paragraphs back, that physicists are going to be flabbergasted if it turns out that these extra conscious states sometimes cause bits of matter to accelerate in your brain. There is one way of endorsing the intuition of distinctness without flouting physical orthodoxy and committing yourself to matter-accelerating conscious forces. You can insist that conscious goings-on are genuinely extra to brain processes, but deny them any causal powers. On this view, brain activity causes an extra conscious 'field', but this field then has no effect on brain activity. What happens in the brain itself is

entirely accounted for by standard physical forces. The extra conscious field 'hovers' above the brain, but the brain runs along its own tracks, as directed by the standard laws of physics.

This position is known as epiphenomenalism. It too has its defenders. David Chalmers takes it seriously in his recent book, *The Conscious Mind* (1996). Chalmers attaches great weight to the intuition that conscious states must be distinct from physical states. Yet he is enough of a scientist to realise that it would fly counter to well-evidenced physical theory to credit these extra conscious states with powers to influence neuronal activity. So he suggests that perhaps they are just epiphenomenal 'danglers', caused by certain kinds of brain activity, but with no power to cause anything themselves. Epiphenomenalism is a cogent position. But it has an obvious drawback. It forces us to deny that our conscious decisions are the causes of our bodily movements. It seems obvious that when I decide to raise my arm, or to go to the pub, my conscious decision is the cause of my limbs moving. But epiphenomenalists must deny this. On the epiphenomenalist view, our conscious mental life 'hovers above' the chains of physical causation that lead from my brain to my bodily movements, without playing any part in them. According to epiphenomenalism, we are like a child in a car with a toy steering wheel, happily twisting it this way and that, blissfully unaware that the actual movements of the car are quite independent of our attempts to steer it.

If we want to avoid this unhappy epiphenomenalist picture, without positing Eccles-style extra mental forces, then we need to return to materialism. That is, we need to deny the intuition that the conscious states are separate from brain states, and insist that decisions and other conscious occurrences are nothing but brain activity, just as heat is nothing but molecular motion. Then, of course, we will have no difficulty understanding how decisions affect behaviour. For if conscious states are brain states, then all we need are the normal physical processes by which brain states cause behaviour. The puzzle about the causal role of conscious states dissolves. It is like the puzzle of how temperature changes manage to influence pressure, given that changes in mean kinetic energy already determine pressure changes. The answer, of course is that temperature *is* mean kinetic energy. Similarly, once we stop thinking in terms of two different states, conscious states and brain states, we don't need to tell any special story to find the conscious states some role in the causal processes. They already have one, in virtue of being one and the same as the brain states that cause behaviour.

Could conscious experiences really be one and the same as brain states? This seems perfectly coherent to me. This materialist position doesn't of course deny that it feels like something to be in some conscious state. It simply identifies this with what it feels like to be in some brain state. That is what it is like, for beings who are in that kind of brain state. (What would you expect it to be like to be in that brain state? To be like nothing? Why?)

### *The Antipathetic Fallacy*

What about the direct intuition that brain states and feelings are quite different in kind? As I said above, I think this is an illusion. We are so close to our own feelings that it is easy to get confused about them. The trouble is that we are able to think about our feelings in a special way—by having them. And this special way of

thinking about feelings makes it difficult for us to see that the things we are thinking about—namely, the feelings—are just the same things as we can think about in other ways—namely, brain states.

It will worth analysing this illusion of distinctness in a bit more detail, for I think that it is responsible for much confusion about consciousness. Let me assume at this stage that materialism is true. This doesn't beg the question. We have already in effect given a strong argument for materialism, by showing that the only alternatives are epiphenomenalism or Eccles-style extra mental forces. The task still facing us is to explain why materialism should seem false, even if it is true. If we can explain this impression, on the assumption that materialism is true, then we will be home.

As a materialist, I maintain that conscious states are identical to brain states, just as heat is identical to molecular motion. But at the same time I recognise that we have two different ways of *thinking about* these states, two different kinds of concepts that refer to these states. By way of analogy, note how the everyday concept of *temperature* and scientific concept of *mean kinetic energy* both refer to the same quantity. Similarly, I say, with the everyday concept *pain*, say, and the physiological concept *nociceptive-specific neuronal activity*. These are two concepts that refer to the same thing.

There is something special about the mind-brain case, however, that makes it very difficult to accept that an everyday concept like *pain* can actually refer to the same thing as a brain state concept like *nociceptive-specific neuronal activity*. Note how mind-brain identity claims contrast with ordinary identity claims in this respect. Once we are shown the evidence, we have no special problem believing ordinary identity claims like temperature = mean kinetic energy. Not so with mind-brain identities. Even after we are shown the arguments for mind-brain identity, we still find it hard to accept that a conscious state can be one and the same as a brain state.

If you ask me, this is because concepts of conscious states pick out their references in a special way. In general, concepts refer to their objects by invoking some description. So when two concepts refer to one thing, this is normally because the two associated descriptions pick out the same thing. For example, The Evening Star and The Morning Star both pick out the same planet, Venus. Similarly temperature and mean kinetic energy can be regarded as two different description of the same quantity, one identifying it in terms of its macroscopic effects, the other in terms of its microscopic constitution. But mind-brain identities work differently. We have special ways of thinking about mental states—we refer to conscious states by using *imagination* or *attention* rather than description. That is, we imaginatively recreate the state in our mind, and then think of it as that *kind* of state (the kind I am recreating now). Alternatively, we focus attentively on the state while we are actually undergoing it, and again think of it as *that* kind of state (the kind I am attending to now).

When somebody refers to conscious states in these special ways, I shall say that they are exercising 'phenomenal concepts'. Such exercises require either that you are actually undergoing the experience referred to, or at least that you are recreating it in your imagination. Note how this means that uses of phenomenal concepts will share their 'what-it's-likeness' with the experiences they refer to. *People who deploy phenomenal concepts—that is, think directly about 'that experience'—will therewith*

*have the feelings involved in the experience.* This is obvious in the case where they refer by attending to an experience while they are having it. But it also holds, to some extent, in the case where they refer to an experience by recreating it imaginatively. Visually imagining a red square is somewhat like actually seeing a red square. It isn't exactly like seeing, of course, but there is an obvious sense in which imagining and seeing are phenomenally similar from the subject's point of view. Similarly, an imagined pain shares some of the phenomenal unpleasantness of a real pain. It doesn't hurt as much, of course, or in the same way, but it can still make you feel queasy, or make you twitch, or make the hairs in your neck stand on end. In Hume's phrase, the imagining is 'a faint copy' of the original impression.

We are now in a position to explain why conscious states should seem intuitively so distinct from brain states, even if they are not. We are misled by the subjective commonality between uses of phenomenal concepts and the experiences thereby referred to. This subjective commonality can easily confuse us when we contemplate identities like *pains = nociceptive-specific neuronal activity*. We focus on the left-hand side, deploy our phenomenal concept of pain (that feeling), and therewith feel something akin to pain. Then we focus on the right-hand side, deploy our concept of *nociceptive-specific neurons*, and feel nothing (or at least nothing in the pain dimension—we may visually imagine axons and dendrites and so on). And so we conclude that the right hand side *leaves out* the feeling of pain itself, the unpleasant what-it's-likeness, and refers only to the distinct physical correlates of pain.

This line of thought is extremely common, both within philosophy and without. When we use our phenomenal concepts, we bring to mind, in a literal sense, something that feels like the experiential state we are thinking about. When we use non-phenomenal concepts, this does not occur. And this makes it seem to us that non-phenomenal concepts cannot possibly refer to the same experiential properties that are picked out by our phenomenal concepts. (Thus consider Colin McGinn's question, on the first page of his *The Problem of Consciousness* (1991), 'How can technicolour phenomenology arise from soggy grey matter?')

However, once we stop to examine it, we can see that this line of thought involves a simple fallacy, indeed a species of Quine's famous use-mention fallacy. There is indeed a sense in which non-phenomenal concepts (like *nociceptive-specific neuronal activity*) do 'leave out' the conscious experiences themselves. They do not *use* such experiences. But it does not follow that they do not *mention* such experiences. After all, most referring terms succeed in denoting their referents without using those referents in the process. There is no reason to suppose that non-phenomenal concepts of experience do not do this too.

Non-phenomenal concepts differ from phenomenal ones in not using the experiences they refer to. This is the sense in which they 'leave out' the experiences. But it does not follow that non-phenomenal concepts differ from phenomenal ones in what they mention. In this referential aspect, which is the one that matters, they need not 'leave out' any element of the experience, not even the 'what-it's-likeness'. There is no reason why we shouldn't be able to refer to this 'what-it's-likeness' using concepts which don't actually give us the feeling. It is only the peculiar fact that some special concepts, our phenomenal concepts, do refer by giving us the feelings which confuses us here.

In my *Thinking about Consciousness* (2002) I dubbed this confusion the ‘antipathetic fallacy’. Ruskin coined the phrase ‘pathetic fallacy’ for the poetic figure of speech that attributes human feelings to nature (‘the deep and gloomy wood’, ‘the shady sadness of a vale’). It seems to me that in the mind-brain case we commit a converse fallacy, and refuse to recognise that conscious feelings inhere in certain parts of nature, namely, the brains of conscious beings.

### *Conclusion*

Let me sum up. If we want to avoid epiphenomenalism and Eccles-style dualism, we need to accept that conscious states are nothing but brain states, just as heat is nothing but molecular motion. Admittedly, this flies in the face of intuition (‘How could technicolour phenomenology arise from soggy grey matter?’) But intuition should not be trusted here. Even if materialism is true, we are easily seduced into thinking it false. This is because concepts of brain states (‘soggy grey matter’), unlike phenomenal concepts (‘technicolour phenomenology’), don’t involve actual experiences. Still, this is no reason for thinking that the brain states themselves don’t involve actual experiences.

Once we expose the antipathetic fallacy, then nothing remains in the way of accepting materialism. This will be good for the study of consciousness. But it will be bad for ‘consciousness studies’. If we accept materialism, we will recognise that there are not going to be any breakthroughs, any crucial discoveries about what ‘causes’ consciousness. That would be like discovering what ‘causes’ life.

Of course there is no such thing to discover. All we can do is classify the different kinds of life, and try better to understand their mechanisms. Similarly with consciousness. We should stop getting excited about the spurious question of what ‘causes’ consciousness. Instead we should settle down to the serious business of classifying kinds of consciousness and exploring their mechanisms.

David Papineau  
*King’s College London*

### *References*

David Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*. (Oxford: Oxford University Press 1996)

John Eccles, *Evolution of the Brain: Creation of the Self*. (London: Routledge 1989.)

Colin McGinn, *The Problem of Consciousness*. (Oxford: Blackwell 1991.)

David Papineau, *Thinking about Consciousness*. (Oxford: Oxford University Press 2002)