

## MIND THE GAP

David Papineau  
King's College London

### 1. Introduction

On the first page of *The Problem of Consciousness* (1991), Colin McGinn asks “How is it possible for conscious states to depend on brain states? How can technicolour phenomenology arise from soggy grey matter?” Many philosophers feel that questions like these pose an unanswerable challenge to physicalism. They argue that there is no way of bridging the “explanatory gap” between the material brain and the lived world of conscious experience (Levine, 1983), and that physicalism about the mind can therefore provide no answer to the “hard problem” of why brains give rise to consciousness (Chalmers, 1996).

I disagree with these philosophers. I think that physicalism has no problem at all in explaining why conscious states go with brain states. This is because I think physicalism is best conceived as a thesis of identity between conscious properties and material properties, and identities need no explaining.

Some of those philosophers who think the explanatory gap cannot be bridged say that they are otherwise persuaded of (or at least open to) physicalism (Nagel, 1974; McGinn, 1991; Strawson, 1994). Why should they say this, if I am right that physicalism leaves no explanatory gap? I shall offer two explanations, one arguing that many who count themselves as physicalists hanker for something they don't need, and the other arguing that many who count themselves as physicalists are often less physicalist than they suppose, and unknowingly slip back into dualist thinking.

In detail, the paper proceeds as follows. In sections 2–5 I shall explain why physicalism about consciousness is best viewed as a claim of identity between conscious properties and physical (or functional) properties. Section 6 will then argue that physicalism then leaves no explanatory mystery, for the simple reasons that identities need no explanation.

I then turn to the question of why so many philosophers, even those who say they are otherwise open to physicalism, nevertheless feel that there is an explanatory gap. In section 7 I point out that many philosophers think that physicalism

ought to explain how certain physical states satisfy descriptions which are associated a priori with our concepts of conscious states, and so are perturbed by the unavailability of such explanations. In response to this worry, I argue that such explanations are indeed unavailable, but that this does not show there is something wrong with physicalism. As I explain in section 8, it simply shows that our normal concepts of conscious states are not associated a priori with any such descriptions. This then leads, in section 9, to a further explanation of why the feeling of a gap is so persistent. Namely, that the structure of our concepts of conscious states makes it very hard for us fully to believe that physicalism is really true.

## 2. How Not to Argue for Physicalism

My first task is to show that physicalism is best conceived as a thesis about property identity. There are various different ways of formulating physicalism. Some writers on physicalism start by comparing these formulations and clarifying the difference between them. This seems to me to come at the topic from the wrong direction. The important question is not how these different formulations compare with each other, but which, if any, we ought to believe.

To address this issue we need to examine the possible arguments in favour of physicalism. If none of these arguments is any good, then it doesn't matter how physicalism is formulated, since no formulation of physicalism will deserve assent. On the other hand, if there are good arguments for physicalism, then they themselves will show us how the doctrine is best formulated, since they will show us which formulations should actually be believed.

I shall look at three styles of argument for physicalism: the argument from intuition, the argument from realization, and the argument from causation. I shall argue that only the last argument works, and that it supports a version of physicalism that identifies conscious properties with physical (or functional) properties.

First, the argument from intuition. Some philosophers seem to think that physicalism can be supported directly by intuition. This is implausible. Those intuitions that are widely shared are impotent to establish physicalism. It is true that nearly everybody nowadays has some intuitions about the correlation between mind and brain. For example, contemporary intuition attests that if we duplicate you, molecule for molecule, in a holocopying machine, then your doppelganger will feel just like you. But this intuition does not suffice for anything worth calling physicalism. To see this, note that the holocopier thought experiment is consistent with epiphenomenalism: perhaps your doppelganger feels like you simply because its brain states causally generate extra conscious states in just the same way as yours do. I take it that this kind of epiphenomenalism clearly doesn't amount to physicalism. Rather, if we are thinking in these terms<sup>1</sup>, we need to show that physical duplicates couldn't *possibly* be mentally different, for only this would show that the mental was somehow constituted by the physical, as opposed to just being correlated with it. But here intuition manifestly fails to deliver. Even dyed-in-the-wool physicalists like myself feel the pull of the intu-

ition that there could be a zombie who is physically just like me but has no feelings—in a possible world, so to speak, which lacks our laws relating conscious states to brain states.

Now for the argument from realization. This style of argument for physicalism begins with the claim that our concepts of conscious states are associated a priori with certain functional descriptions. It then appeals to empirical science to show that in this world these functional descriptions are in fact filled by physical states<sup>2</sup>. The trouble with this strategy, however, is that a priori conceptual analysis does not in fact seem to reveal associations between our concepts of conscious states and any functional descriptions. I shall return to this point at some length in sections 7–8.

### 3. The Causal Argument for Mind-Brain Identity

Physicalists need not lose heart. There is another kind of argument which does establish physicalism. It simply points out that, as a matter of empirical fact, conscious states and physical states always seem to appear at the same place in the causal scheme of things, and concludes that they must therefore be identical. Let me lay out this argument more formally. As a first premise, take:

- (1) Conscious mental occurrences have physical effects.<sup>3</sup>

Now add in the premise (“the completeness of physics” henceforth) that:

- (2) All physical effects have sufficient physical causes.<sup>4</sup>

At first sight, these two premises suggest that a certain range of physical effects (think of your physical behaviour) have two distinct sufficient causes, one involving a conscious state (your pain, say), and the other purely physical (your C-fibres firing<sup>5</sup>).

Now, some events are indeed overdetermined in this way, like the death of the 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 my arm would still have moved even if I hadn’t felt the pain (because my C-fibres were firing), or, alternatively, even if my C-fibres hadn’t fired (because I felt a pain). So let us add the further premise:

- (3) The physical effects of conscious causes aren’t always overdetermined.

Now, there seems an obvious and plausible way to make (1)–(3) consistent. We need to identify the conscious occurrences mentioned in (1) with (parts of) the physical causes mentioned in (2). This then allows us to uphold both (1) and

(2), and yet avoid the implication of overdetermination, since (1) and (2) no longer imply distinct causes.

Exactly what is involved in identifying conscious causes with physical causes? In the first instance, it depends on how you think of the relation of causation. I think of these as facts. In the present context of argument, it will come to the same thing if you think of them as states or as “Kim-events”, that is, as instantiations of properties by particulars. For, on all these views, the identification of conscious and physical causes requires that the conscious property (of being in pain, say) is itself identical with the physical property (having C-fibres firing), since the relevant facts, or states, or Kim-events, won't be identical if these properties aren't.

What if you don't think of the relation of causation in any of the above ways, but as basic particulars, as Donald Davidson does? 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 only requires that the relevant causal and physical properties are instantiated in the same particular, not that the properties themselves are identical. Still, we can rephrase the argument. Let us take premise (1') to be the claim that all conscious events cause some physical events in virtue of their conscious properties; premise (2') says that all physical events are caused by prior physical events in virtue of the latter's physical properties; and premise (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.

Property identity claims of this sort will provide the crucial premise for the main argument of this paper. But before proceeding I would like to elaborate a bit further on the causal argument. First, in the next section, I shall briefly consider the strength of the argument's premises. Then, in section 5, I shall discuss whether the argument can be squared with the apparent variable physical realization of conscious properties. Both these sections are of the nature of digressions. Readers who are primarily interested in the explanatory gap may wish to skip ahead to section 6.

#### 4. Rejecting the Premises of the Causal Argument

All three premises of the causal argument can be denied, but each of the resulting positions seems less attractive than accepting the identity of conscious and physical properties.

You can reject premise (1) and embrace epiphenomenalism or psychophysical parallelism. I take it that we would prefer to avoid these views if we can. We don't seem to find mechanisms involving causal danglers or brute parallels in other parts of the natural world. This gives us inductive reason not to expect them at the mind-brain interface either.

You can reject premise (2), and deny that physics is causally complete. This suggestion raises a number of interesting issues, including questions about the

evidence which has persuaded scientists, over the last one hundred and fifty years, of the completeness of physics. Still, I think that in the end (2) must be upheld. The question you need to ask yourself is whether particles of matter (sub-atomic particles in your brain, maybe) are ever caused to accelerate by conscious causes alone, in the absence of any other forces. In effect, do we need to include purely conscious causes alongside gravity, the electroweak force, and so on, in the category of fundamental forces? This is not an incoherent thought, and it has its defenders (e.g. Menzies, 1988), but I take it that it runs counter to a wealth of empirical evidence. (Your colleagues in the Physics Department would certainly be very interested if such a force could be shown to exist.)<sup>6</sup>

You might even reject premise (3), and accept that the physical effects of conscious causes also have distinct physical causes. On this “belt and braces” view, there are always two different causes to make your arm move. Gabriel Segal and Elliott Sober (1991) and D.H. Mellor (1995, pp. 103–5) seem ready to embrace this view. They observe that the distinct mental and physical causes may themselves be strongly counterfactually dependent on each other, and that this may block the unhappy inference that your arm would still have moved even if you hadn’t felt a pain, since without the pain your neurones wouldn’t have fired either. Still, these philosophers have their work cut out to explain exactly why the two causes should be so strongly counterfactually dependent. And, in any case, as with danglers and parallels, we don’t seem to find belt-and-braces mechanisms elsewhere in the natural world, which once more gives us inductive reason to be suspicious of postulating them here.

## 5. Physical or Functional?

Let me turn now to the question of whether the causal argument can be squared with the apparent variable physical realizability of conscious properties. The conclusion of the argument, as presented so far, is that conscious properties are identical to strictly physical ones. However, it seems highly plausible that beings who have nothing physical in common can nevertheless share conscious properties. For example, it is plausible that octopuses, who lack C-fibres, can nevertheless instantiate the property of being in pain. So the property of being in pain cannot be identical with having C-fibres firing. (All right, C-fibres themselves are arguably not strictly physical. But they are physical enough to make the point.)

The standard move at this point is to reject the identity of pain with C-fibre firings, and instead identify it with whatever second-order (“functional”, “role”) property is common to pained humans and octopuses. (A first pass might be the property-of-having-some-physical-property-which-arises-from-damage-and-gives-rise-to-avoidance-behaviour.)

If we take this functionalist option, however, it is not clear that we are still respecting the premises of the causal argument outlined in the last section. In particular, it is not clear we are respecting premise (1). For, on the functionalist view, conscious properties are no longer identical with the strictly physical prop-

erties involved in the causes of behavioural effects, but rather identical with certain role properties which are in turn realized by those physical causes. So the conscious states don't strictly cause those behavioural effects, at least not in the same sense as the physical states cause them.

Still, perhaps we can re-jig the causal argument to make it deliver the functionalist conclusion. Suppose we begin again, and assume only that mental states cause physical effects in the weaker sense that either they cause them directly or they have realizers that cause them directly. That is, we might read "cause" in premise (1) in a generous sense, which allows a state to "cause" in virtue of having a realizer which causes. If we do this, then (1)–(3) will no longer require us to identify conscious states with strictly physical states, but it will still give us an argument for identifying them with second-order states which are physically realized. For unless we suppose this identification, we won't even be able to satisfy the weaker version of premise (1), and will be in danger once more of having our behavioural effects overdetermined by two ontologically quite unrelated causes.<sup>7</sup>

It is important not to confuse this re-jigged version of the causal argument with the realization argument mentioned earlier in section 2. The earlier idea was that we could *base* an argument for physicalism on an *a priori* connection between concepts of conscious states and functional descriptions; and the difficulty with this, as I pointed out, is that there seems to be no such a priori conceptual connection. The current suggestion, by contrast, takes the identity of conscious properties and role properties to be the *conclusion* of an argument with *a posteriori* premises. It does not require any a priori connection between conscious properties and functional descriptions. Rather, it starts with (the re-jigged) premises (1)–(3), all still read as empirical claims, and concludes that, if conscious properties are not strictly identical with physical properties, then they must at least be identical with role-properties-which-are-physically-realized, otherwise we won't be able to see them as causing their physical effects. Nothing is here being assumed about our *concepts* of conscious properties, nor a fortiori about their a priori connection with functional descriptions. The conclusion is simply that, whatever the precise nature of our concepts of conscious properties (a topic to which I shall return), they must in fact *refer* to physically realized functional properties.

It won't matter for the rest of this paper whether we take the causal argument to establish that conscious properties are a posteriori identical to strictly *physical* properties, or to physically realized *functional* properties.<sup>8</sup> Whichever way we go, we will still have an identity between conscious and other properties, which is what I need for the main argument of this paper. The crucial premise for this argument is that physicalism is a claim of property identity. The argument will work equally well whether the identity is with strictly physical or with physically realized role properties. Accordingly in what follows I shall use the term "material property" as a coverall for both strictly physical properties and physically-realized role properties.

## 6. Identities Need no Explaining

If conscious properties are identical to material properties, then I say there is no mystery of why material properties “give rise” to conscious properties. This is because identities need no explaining. If the “two” properties are one, then the material property doesn’t “give rise” to the conscious property—it *is* the conscious property. And if it is, then there is no mystery of why it is what it is.

An analogy will help to make the point clear. Suppose you don’t know that Tony Curtis and Bernie Schwartz are the same person. Then you are told that they are identical. Now, this might well prompt you to ask for an explanation of what shows they are identical (and the answer, presumably, would be that they always appear at the same place in the causal scheme of things). But it would make no sense for you then to ask for a further explanation of *why* they are identical. “Why does Tony Curtis = Bernie Schwartz?” seems almost ill-formed. If they are one, then they are. That single person couldn’t possibly have been two people.

This example involves an identity between spatio-temporal particulars. But the same goes for identities between universals. Suppose you don’t know that water = H<sub>2</sub>O. Then you are told that they are in fact the same stuff. You can sensibly ask for an explanation of what shows they are the same stuff (and the answer, again, will be that they appear at the same place in the casual scheme of things). But it does not make good sense to ask in addition *why* they are the same stuff, to ask for some further explanation of why water is H<sub>2</sub>O. If they are the same stuff, they are, and that’s it. (Cf. Block, 1978.)

Exactly the same point, I say, applies to the identity of conscious and material properties. We can ask for an explanation of what *shows* that pains are identical to the firing of C-fibres (once more, they appear at the same place in the causal scheme of things). But it makes no sense then to ask for a further explanation of *why* pains arise from C-fibre firings. If they are C-fibre firings, they don’t “arise” from them, they are them, and that’s it.

Perhaps it is worth making clear that I certainly do not want to deny the “what-it’s-likeness” of conscious occurrences. To say that pains are identical with C-fibres firing is not to deny that it is like something to be in pain. Rather, it is to affirm that it is like something to have your C-fibres firing. Indeed that is precisely what we have found out, via the causal argument in Section 3—feeling a pain is having your C-fibres firing. And given this, there is then no further mystery of why firing C-fibres should feel like that. The “two” states are the same, and that’s it.

## 7. Role-Filling Explanations

I suspect that many readers will feel quite unsatisfied by the argument of the last section. Even those who are persuaded, by the causal argument from sections 3–5, that conscious properties are identical to material properties, and who ac-

cept that in general identities need no explanation, may well feel that there is something special about the mind-brain case that calls for further explanation.

In the rest of this paper I shall try to respond to this feeling. I shall offer two different accounts of why someone might still feel that consciousness needs explaining, even given the arguments I have rehearsed so far. In this section I shall look at a difference between the mind-brain case and most other identity claims in science. A number of philosophers of consciousness attach great weight to this difference, but I myself do not think it gets to the heart of the intuitive feeling that consciousness is inexplicable. Even so, it will be useful to examine the issue, as it will open the way to a more satisfying account of the intuition of inexplicability, which I shall develop in the final two sections of this paper.

By way of introduction to this first explanation of the appearance of inexplicability, note that, despite the arguments of the last section, there is clearly one sense in which some identities can be explained. Take the claim that Tony Blair = the British prime minister. Obviously it makes sense to ask why Tony Blair is the British prime minister, and plenty of reasonable answers might be given, such as that he won the most votes in the election, or that the Labour party thought he was the leader most likely to defeat the Conservatives, or... (where the exact answer you give will depend, as with all explanations, on the interests and epistemic needs of your audience).

It is clear enough what is going on here. When an identity claim is flanked on one side by a description, we can ask for an explanation of why the entity in question uniquely satisfies that description, and often enough we are in a position to provide a satisfying answer. This does not invalidate the point made in the last section. We are not here explaining why the entity is itself, so to speak. That still needs no explanation. Rather, we are explaining why it satisfies some description, having been pointed to this question by the appearance of the description in our identity claim.

Something like this kind of explanation can be associated with standard scientific identity claims. Take the claim that water is  $H_2O$ . If we understand the term “water” as in some sense a priori equivalent to “the familiar liquid which is colourless, odourless and tasteless”, then we can sensibly ask why  $H_2O$  is water, and read this as a request for an explanation of why  $H_2O$  is colourless, odourless and tasteless, a request which can in principle be answered by reference to the physical chemistry of  $H_2O$ . And, in general, if it is claimed that some theoretical entity discovered by science is identical with some everyday stuff or quantity (lightning is electrical discharge, heat is molecular motion), there would seem to be room for an explanation of why the theoretical entity in question satisfies the descriptions associated a priori with the relevant everyday concept.

Some philosophers feel we ought to be able to do something similar in the mind-brain case. We ought to be able to explain why certain physical states satisfy the descriptions associated a priori with our everyday concepts of conscious states. But, they then point out, we don't seem to be able to do this. Thus they conclude that there is something about the relation between conscious and physical states that we ought to be able to explain, but can't.

This lack will seem particularly pressing if you think that the canonical argument for physicalism is the argument from realization mentioned in section 2. For then you will view the absence of the desired explanations, not just as a disturbing explanatory gap, but rather as an obstacle blocking any epistemological access to physicalism. (Cf. Levine, 1993; Chalmers, 1996, ch. 2.)

I agree that we lack explanations of the kind being asked for here. However, I don't agree that this reflects badly on physicalism. For the reason we lack such explanations is simply that there are no descriptions associated a priori with our everyday concepts of conscious states. So it is scarcely surprising, and no argument against physicalism, that we can't show why certain physical states fulfil the a priori functional requirements for being a pain, say. If there are no such a priori requirements, then there is no issue of why certain physical states satisfy them.

Nor should we conclude, in the absence such explanations, that we have no reason to believe physicalism in the first place. This negative conclusion would follow if the only epistemological route to physicalism were via the realization argument mentioned in section 2. But this is not so. The causal argument provides a far better route to physicalism, which does not assume any a priori associations between our everyday concepts of conscious states and any descriptions.

From this point of view, scientific identities like  $\text{water} = \text{H}_2\text{O}$  provide a bad model for the physicalist identification of mental with material properties. In the scientific case we can indeed show how  $\text{H}_2\text{O}$  satisfies certain descriptions associated a priori with "water", and use this as a route to the conclusion that water is  $\text{H}_2\text{O}$ . However, the assumption that all physicalist identifications must conform to this pattern does physicalism a disservice. This pattern of argument is fine when our pre-theoretic concepts are associated a priori with functional descriptions, but there is no need to insist on this form of identification when our concepts are not.<sup>9</sup>

Some terminology of David Chalmers' will help to make it clear why there are no functional descriptions associated a priori with our everyday concepts of conscious states. Perhaps the word "pain" does have some analytic connections with certain descriptions—"deriving from physical injury", "generating avoidance behaviour", and so on. But, in this use, says Chalmers, "pain" just expresses a "psychological" concept. However, we also have a "phenomenal" concept of pain—as something with a certain what-its-likeness. And this phenomenal concept of pain is quite different from the psychological concept, as is shown by the fact that there is no a priori reason to suppose that their referents will always be found together. If we now know that the feeling (the unpleasant one) is generally present in individuals who are in the psychological state, it is on the basis of empirical evidence. As far as reflection on our concepts goes, it seems possible that some other feeling, or none at all, should generally accompany things like bodily damage and writhing around.

I agree about all this, and am happy to concede that the important question is what our *phenomenal* concepts refer to. Chalmers is quite right to insist that the crucial concepts at issue are our phenomenal concepts. They are the ones that

make people feel so puzzled about the mind-brain relation. Still, even if phenomenal concepts, unlike psychological concepts, have no a priori connection with any descriptions, it doesn't follow that they don't refer to material properties.<sup>10</sup> For, to repeat, my preferred causal route to this conclusion doesn't presuppose any a priori connection between our phenomenal properties and any descriptions.

## 8. Concepts of Conscious States

How do our phenomenal concepts refer, if not by a priori association with descriptions? This is a good question, which puts the physicalist under some pressure. Some philosophers hold that any satisfactory account of the distinctive workings of our phenomenal concepts will have to mention *phenomenal properties* which are distinct from any material properties. In support of this view, these philosophers point out that it is generally not possible to possess a phenomenal concept unless you have had the corresponding conscious experience. (Jackson, 1982, 1986.) Somebody who has never seen anything red cannot deploy a phenomenal concept of red visual experience, even if they already possess concepts of all the physical and functional aspects of red experiences. This suggests that the phenomenal concept somehow derives from first-hand acquaintance with some distinctive phenomenal property—the *feel* of red experiences.

However, it is not essential that we account for the workings of phenomenal concepts in this way. Suppose that the deployment of phenomenal concepts involve a kind of *simulation* of the experiences they refer to. After all, there is now plenty of evidence, in the form of various kinds of brain scans, to show that, when you use a phenomenal concept, you activate (some of) the same parts of the brain as are involved in the original experience itself.

This simulational model suggests an alternative physicalist account of the acquisition and subsequent deployment of phenomenal concepts, an account which makes no reference to any supposed acquaintance with distinctive phenomenal properties.

Let me take the *acquisition* and subsequent *deployment* of phenomenal concepts in turn. The simulational model of phenomenal concepts offers an obvious explanation of why we can only *acquire* a phenomenal concept after previously having had the corresponding experience: namely, that human beings can only “re-activate” those patterns of neurones which have previously been activated by some independent source. There is no reference in this story to any acquaintance with distinctive phenomenal properties. The idea is simply that, as a matter of contingent fact, human beings only acquire the ability to “switch on” the relevant parts of their brains once those parts have previously been stimulated by some external source.

Let me now consider the subsequent *deployment* of phenomenal concepts. There are two obvious ways in which we deploy phenomenal concepts, once we have acquired them. First, we can use them to classify (directly, and without appeal to behavioural or physical evidence) our current experiences. Second, we can use them to phenomenally imagine such experiences. Neither of these uses of

phenomenal concepts demands explanation in terms of distinctive phenomenal properties. When we classify our current experiences directly, this needn't involve our noting introspectively that they have some distinctive phenomenal property. Rather, we may simply have some mechanism which compares the current state of our brain with a "template" created by previous experiences, and then directly "triggers" the verdict "that experience again". (Cf. Lewis, 1988.) Similarly, when I imagine someone feeling a pain, say, or seeing red, this needn't involve attributing them distinctive phenomenal properties. Rather, I may simply "re-activate" (some of) those parts of my brain that are activated by the experience itself, and then "project" that state onto whomever I am thinking about, with the thought "so-and-so has *this* type of experience". (Cf. Loar, 1990.)

If we accept this simulational model of phenomenal concepts, it is clear why there is no question of explaining how certain physical states satisfy the descriptions associated with our phenomenal concepts. There are no such descriptions. Phenomenal concepts refer via a species of simulation, without invoking any descriptions. This is why "H<sub>2</sub>O = water" is not a good analogy for "C-fibres firing = pain" Since the term "pain" expresses a phenomenal concept, it is not associated a priori with any descriptions, and so there is no possibility of showing how associated descriptions pick out a physical state, as we can with "water".<sup>11</sup>

The original example of "Tony Curtis = Bernie Schwartz" turns out to provide the better analogy. Proper names are not normally associated with canonical descriptions. This is why our minds go blank when asked to explain *why* Tony Curtis is Bernie Schwartz. The same barrier, I contend, prevents us explaining why C-fibres firing is pain.

## 9. Why We Still Feel There is a Gap

In my view, the technicalities of the last section have little to do with the real reason why people hanker for some further explanation of the mind-brain connection. The last section pointed out, in effect, that you might feel something was missing if you were presented with a "Tony Curtis = Bernie Schwartz" identity, and somehow read it as a "Tony Blair = the British prime minister" identity. Similarly in the mind-brain case, so the suggestion went, you might feel the need for something more, even after you embrace the identity of pain with some material property, if you mistake the phenomenal concept of pain for a concept that refers in virtue of its association with descriptions.

Well, maybe so. But that all seems a bit high-falutin' to me, and I doubt it really has much to do with the persistent feeling that there is something mystifying about the mind-brain relation. My own diagnosis of the feeling of mystery is that most people are simply not prepared to accept that phenomenal concepts refer to material properties in the first place. Even when faced with the strong (causal) arguments that phenomenal concepts must refer to material properties, most people remain convinced that they refer to distinct conscious properties. And then of course they do have something to explain. For once they suppose that phenomenal concepts refer to conscious properties, distinct from material prop-

erties, then they are faced with the extremely hard conundrum of why certain material properties should always “give rise to” these special conscious properties.

So in my view the problem of explaining the appearance of an explanatory gap reduces to the problem of explaining why most people are so resistant to accepting that conscious properties are identical with material properties. Why are people so disinclined to accept this, even in the face of evidence that these properties are always instantiated in the same places at the same time, that they seem to play exactly the same roles in the causal scheme of things, and so on?

Here we have a disanalogy with identities of the “Tony Curtis = Bernie Schwartz” type. Once we are shown the evidence for this identity, we don’t go on asking why Bernie Schwartz is sure to go everywhere Tony Curtis goes. But even after we are shown that conscious properties make no separate appearance in the causal scheme of things, we continue to feel they are distinct properties, and so continue to wonder why they always appear alongside certain material properties.

Here is my explanation of why people are so disinclined to accept mind-brain identity. It relates to the analysis of phenomenal concepts given at the end of the last section. Phenomenal concepts may be similar to proper names in not invoking descriptions, but they are also dissimilar in that they refer by simulating their referents. This peculiar feature of phenomenal concepts gives rise to a powerful illusion of mind-brain distinctness. Elsewhere I have called this illusion “the antipathetic fallacy” (Papineau, 1993a, 1993b, 1995.) I believe that this fallacy is the real reason why so many people think the mind-brain relation mysterious.

The antipathetic fallacy arises because some, if not all, imaginative uses of phenomenal concepts share their “what-its-likeness” with the experiences they refer to. Visual imagination provides the clearest examples. Imagining seeing a red square is somewhat like actually seeing a red square. Visually imagining 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. In Hume’s phrase, the imagining is “a faint copy” of the original visual impression. (This is perhaps unsurprising, given the point, noted in the last section, that imaginings activate some of the same parts of the brain as the original experiences.)

Nor is the phenomenon restricted to the visual realm. 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. Again, imagining tasting chocolate feels akin to actually tasting chocolate. Even if it’s not as nice, it can still make your mouth water.

This subjective commonality between the imaginative deployment of phenomenal concepts and the experiences they refer to can easily confuse us when we contemplate identities like pains = C-fibres firing. We focus on the left-hand side, deploy our phenomenal concept of pain (*that feeling*), and feel a teeny bit twingy. Then we focus on the right-hand side, deploy our concept of C-fibres firing, and feel nothing (or at least nothing in the pain dimension—we may visually imagine nerve cells and so on). And so we conclude that the right hand side

leaves out the feeling of pain itself, the unpleasant what-its-likeness, and refers only to the distinct physical correlates of pain.

I think that this line of thought is extremely common, both within philosophy and without. When we use our phenomenal concepts imaginatively, we bring to mind, in a literal sense, an instance of the experiential property 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 denote the same experiential properties that are picked out by our phenomenal concepts. (Thus consider McGinn, with my italics: “How can *technicolour phenomenology* arise from *soggy grey matter*?”)

However, this line of thought involves a simple fallacy, indeed a species of the use-mention fallacy. There is indeed a sense in which non-phenomenal concepts (like *C-fibres firing*) 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.

This then is my explanation of the apparent explanatory gap. Many philosophers, including some who profess physicalism, are seduced by the antipathetic fallacy. They note that when you think in terms of non-phenomenal concepts (“*soggy grey matter*”) you don’t have the feelings you have when you think with phenomenal concepts (“*technicolour phenomenology*”). This then persuades them, fallaciously, that the feelings are something extra to any material properties (and therewith, by my lights, persuades them out of their physicalism). And then of course they face the problem of explaining why the mysterious extra feelings should arise when they do.

The solution, of course, is to resist the fallacy. The feelings aren’t something extra. Having feelings is just what it is to be in certain material states, when we are in those states. Once we fully accept this, and stop sliding back into dualism, then we can stop looking for any explanation of why those states are what they are.<sup>12,13</sup>

## Notes

1. That is, in terms of supervenience. I myself think that notions of supervenience tend to muddy the waters of physicalism. Still, if you do think in terms of supervenience, the

- present point is that physicalism should be formulated as supervenience across all logically, or metaphysically, or at least all physically possible worlds, given that mere co-variation across all *nominally* possible worlds doesn't rule out epiphenomenalism. (Helen Steward and Gene Witmer helped me to see this point clearly.)
2. The classic source for this form of argument is Lewis, 1966. Note that it doesn't immediately follow from this form of physicalism (mental roles are in fact filled by physical states) that the mental will metaphysically supervene on the physical (think of a world which differs mentally from ours, not because it is *physically* different, but because some extra angel stuff also fills mental roles there). True, you can with some effort force this version of physicalism into the mould of supervenience (Chalmers, 1996: 38–41), but it's not clear why you should bother. The doctrine that all mental roles are realized by physical states in the actual world is a clear enough version of physicalism in its own right.
  3. Note that (1) doesn't claim that our *concepts* of conscious *types* are a priori equivalent to the concepts of specific causal roles, as in the Lewis-style argument for identity. It is the far less contentious claim that, as a matter of a posteriori fact, particular conscious states have particular physical effects.
  4. What about quantum indeterminacy? A stricter version of (2) would say that the chances of physical effects are always fixed by sufficient physical causes, and 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 what follows.
  5. Let us follow philosophical convention and take "C-fibres firing" as an approximation for the physical state which is present in humans who are in pain.
  6. In case you think that the vagueness of the term "physics" undermines premise (2), note that none of the substantial points in the above paragraph (those following "The question you need to ask...") uses the term "physics". (For more on this issue, see Crane, 1991, and Papineau, 1991.)
  7. Won't we have two causes anyway, namely, (a) the role property with which we are now identifying the conscious property, and (b) the physical property which directly causes the behavioural result? Well, we might in a sense have two "causes", but they won't *overdetermine* the result, if the role property is present only in virtue of the physical property's presence. (Note that in this case the behavioural result *wouldn't* still have occurred if the physical property had been absent, for then the role property would have been absent too; and similarly, if the role property had been absent in any particular case, so would the physical property have been absent.)
  8. Myself, I feel pulled both ways. There are attractions in following David Lewis and identifying pains with physical types. "Pain" then refers to different properties in humans and octopuses, which properties can therefore be identified with strictly physical and therefore fully efficacious properties (Lewis, 1980). But it seems a bit odd to deny that different types of being can share conscious properties. So perhaps we should stand by variable realizability after all, and recognize that conscious states, as second-order states, do not cause physical effects in the same strict sense that physical states do, but only in the sense that they are realized by physical states which are fully causally efficacious. For further discussion of the causal argument, see Crane (1995).
  9. Indeed it is worth noting that, while we can nowadays sketch quantum mechanical explanantions of why H<sub>2</sub>O satisfies the requirements of colourlessness, tastelessness, and so on, scientists became convinced that water is H<sub>2</sub>O long before they could do this,

- in the first half of the nineteenth century, on grounds that were more like our argument from causation than any argument from realization.
10. Indeed, a phenomenal concept may well refer to just that second-order property which mirrors the descriptions involved in the associated “psychological concept” (such as: the property of having some property which responds to damage and causes avoidance behaviour). This would fit with the “re-jigged” version of the causal argument developed in section 5 above, which concluded that our concepts of conscious states in fact refer to physically-realized second-order states. Note that, if we do take the causal argument this way, then we might wish to explain *how* the relevant second-order property is physically realized (the pain-role is realized by C-fibres in humans, say, and perhaps by other physical states in other beings). But *this* explanation, as Chalmers rightly points out, won’t itself explain why the referent of our *phenomenal* concept should accompany the physical realizer, since this explanation does nothing in itself to tie our phenomenal concept to the role property whose realization is being explained. To understand why the referent of our phenomenal concept should always accompany the physical realizer, we *also* need the causal argument, to tell us that the phenomenal concept in fact refers to the role property which is filled by this realizer.
  11. For a rather different account of why the structure of our phenomenal concepts precludes “realization explanations”, see Sturgeon (1994).
  12. It may have occurred to some readers that the argument in the latter sections of this paper mirrors one possible line of response to Kripke’s objection to physicalism. This correspondence is only to be expected. Where Kripke challenges physicalists to account for the appearance of mind-brain contingency, given their view that the relation is necessary, I have accounted for the appearance of an explanatory mind-brain gap, given my view that there is no such gap. The connection between the two issues is that it is precisely contingencies which need explaining, by contrast with genuine identities, which are necessary, and so in no need of explanation. (For more on the relevant response to Kripke, see Papineau, 1993b, p. 118.)
  13. I would like to thank Tim Crane, E.J. Lowe, Michael Pauen, Gabriel Segal and Scott Sturgeon for comments on a draft of this paper. I would also like to thank Rick Lewis for the title.

## References

- Block, N. (1978). “Reductionism: Philosophical Analysis,” in *Encyclopedia of Bioethics*. London: Macmillan.
- Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.
- Crane, T. (1991). “Why Indeed?” *Analysis* 51: 32–7.
- Crane, T. (1995). “The Mental Causation Debate,” *Aristotelian Society Supplementary Volume* 69: 211–36.
- Jackson, F. (1982). “Epiphenomenal Qualia,” *Philosophical Quarterly* 32: 127–36.
- Jackson, F. (1986). “What Mary Didn’t Know,” *Journal of Philosophy* 83: 291–5.
- Levine, J. (1983). “Materialism and Qualia: The Explanatory Gap,” *Pacific Philosophical Quarterly* 64: 354–61.
- Levine, J. (1993). “On Leaving Out What It’s Like”, in M. Davies and G. Humphreys (eds) *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
- Lewis, D. (1966). “An Argument for the Identity Theory,” *Journal of Philosophy* 63: 17–25.
- Lewis, D. (1980). “Mad Pain and Martian Pain,” in Block, N. (ed) *Readings in the Philosophy of Psychology vol 1*. Cambridge, Mass: Harvard University Press.

- Lewis, D. (1988). "What Experience Teaches," *Proceedings of the Russellian Society of Sydney University*; reprinted in Lycan, W. (ed) *Mind and Cognition*. Oxford: Basil Blackwell.
- Loar, B. (1990). "Phenomenal States," *Philosophical Perspectives* 4: 81–108.
- McGinn, C. (1991). *The Problem of Consciousness*. Oxford: Basil Blackwell.
- Mellor, D. (1995). *The Facts of Causation*. London: Routledge.
- Menzies, P. (1988). "Against Causal Reductionism," *Mind* 97: 551–74.
- Nagel, T. (1974). "What is it Like to be a Bat?" *Philosophical Review* 4: 435–50.
- Papineau, D. (1991). "The Reason Why: Response to Crane," *Analysis* 51: 37–40.
- Papineau, D. (1993a). "Physicalism, Consciousness, and the Antipathetic Fallacy," *Australasian Journal of Philosophy* 71: 169–183.
- Papineau, D. (1993b). *Philosophical Naturalism*. Oxford: Basil Blackwell.
- Papineau, D. (1995). "The Antipathetic Fallacy and the Boundaries of Consciousness," in Metzinger, T. (ed) *Conscious Experience*. Paderborn: Schöningh.
- Segal, G. and Sober, E. (1991). "The Causal Efficacy of Content," *Philosophical Studies* 63: 1–30.
- Strawson, G. (1994). *Mental Reality*. Cambridge, Mass.: MIT Press.
- Sturgeon, S. (1994). "The Epistemic View of Subjectivity," *Journal of Philosophy* 91: 221–235.