

David Papineau

## *Comments on Galen Strawson*

*'Realistic Monism:  
Why Physicalism Entails Panpsychism'*

### **1. Straightforward Physicalism**

Galen Strawson (2006) thinks it is 'obviously' false that 'the terms of physics can fully capture the nature or essence of experience' (p. 4). He also describes this view as 'crazy' (p. 7). I think that he has been carried away by first impressions. It is certainly true that 'physicalism', as he dubs this view, is strongly counterintuitive. But at the same time there are compelling arguments in its favour. I think that these arguments are sound and that the contrary intuitions are misbegotten.

In the first two sections of my remarks I would like to spend a little time defending physicalism, or 'straightforward' physicalism, as I shall call it ('S' for 'straightforward', if you like). I realize that the main topic of Strawson's paper is panpsychism rather than his rejection of straightforward physicalism. But the latter is relevant as his arguments for panpsychism depend on his rejection of straightforward physicalism, in ways I shall explain below.

Let me first explain how I understand straightforward physicalism. I don't suppose that current science has yet established, for any specific phenomenal kind M and any specific physical kind P, that  $M = P$ . Our understanding of the brain is as yet too fragmentary. Indeed it is possible that our access to such identities will require significant advances in brain science. Even so, I see no reason to think that such advances need to take us beyond physical theories of the same general sort as we already have. So I do suppose that every phenomenal kind M is identical to *some* P that is generally similar to the kinds currently

*Journal of Consciousness Studies*, **13**, No. 10–11, 2006, pp. 100–109

recognized by the physical sciences.<sup>1</sup> Moreover, I suppose that when we have established such  $M = P$  identities, then we will therewith have ‘fully captured the nature or essence of experience’ in physical terms, in that the relevant physical term will refer to nothing other than the phenomenal kind  $M$ .

I shall not here rehearse the arguments in favour of this straightforward physicalism. These are well-known, and derive from familiar causal-explanatory considerations (cf. Papineau, 2002, ch. 1). Rather, the question we need to ask is what countervailing reasons might nevertheless block these arguments. Strawson is by no means alone among contemporary thinkers in judging that the familiar causal-explanatory arguments are somehow defeated by the special nature of consciousness, and that therefore we must find some alternative to straightforward physicalism (such as epiphenomenalism or Strawson’s own ‘real physicalism’). But what exactly is it about consciousness that is supposed to prevent the causal-explanatory arguments from establishing their straightforward conclusion?

One possible thought is that it is simply a conceptual contradiction to suppose that a conscious state might be one and the same as a physical state. Doesn’t our concept of a physical state definitionally exclude any conscious element, and vice versa? Isn’t there something contradictory in supposing that a being with only physical properties could have conscious experience?

But this doesn’t strike me as contradictory at all. As far as our concepts go, it seems to me completely open what it would feel like to be a purely physical being with firing C-fibres, say. What would you expect that to feel like? Why suppose it must feel like nothing? That’s one possibility, but conceptually it seems just as possible that it would feel like being light-headed, or indeed feel like being in pain. Our concepts are not going to decide this issue.

Perhaps the problem is not that our concepts of physical states conceptually exclude any conscious element, but that they *don’t conceptually guarantee* any.<sup>2</sup> We certainly can’t tell straight off that a physical being with C-fibres firing will be in pain. (As is now conventional, let me use C-fibres as a place-holder for whichever physical state is in fact correlated with pain.) At best, the equation of pains with C-fibre firings will be the result of empirical investigation.

[1] In these remarks I will use ‘physical’ broadly, and not just to refer to kinds studied in physics departments. So it should be understood as including physically realized role states along with strictly physical states.

[2] This lack of a conceptual guarantee is the ‘standard argument’ against physicalism that Strawson cites in footnote 4, referring to Strawson (1994), pp. 62–5.

Still, why is that any reason to deny that pains are one and the same as C-fibre firings? Many true identities are only established a posteriori, such as that table salt is sodium chloride, or that light is electromagnetic radiation, or that Cary Grant is Archie Leach. But that is no reason to deny, say, that ‘the term *sodium chloride* fully captures the nature and essence of table salt’. So why not similarly allow that the term *C-fibres firing* fully captures the nature and essence of pain?

Maybe the underlying thought is that identities and other necessities only appear a posteriori when we are thinking about them in a second-hand way that does not *acquaint* us directly with the entities involved. When we think about NaCl as *table salt*, we are thinking about it indirectly, as *the stuff, whatever it is, that is left when sea water evaporates, is good for flavouring food, etc.* That is why we can’t tell straight off that it is NaCl. But our concept of *pain* surely doesn’t similarly pick out its referent indirectly. So if pain = C-fibre firings were true, it should be a priori, given that we think about pain (and C-fibres) directly, and not at second-hand. But of course pain = C-fibre firings is not a priori. So it can’t be true. Or so at least this story goes.

This line of thought has been gaining currency over the past decade (Chalmers, 1996; Jackson, 1998). But the notion of direct acquaintance on which it rests strikes me as highly suspicious. It assumes some mode of thought where objects become completely transparent to the mind and all their essential properties are thereby laid bare. It is hard not to see this as inspired by some misplaced visual model, in which we are able to peer in at some immaculately illuminated scene. I myself doubt that there is any such mode of thought. No doubt there are ways of thinking of things that make certain essential properties a priori knowable. But I take such a priori knowledge to derive from (possibly implicit) compositionality in the relevant modes of thinking, and so not to be associated with the most basic ways in which thought makes contact with reality.

When it comes to these basic points of contact, I find it hard to take seriously any alternative to the assumption that our atomic concepts are related to reality by facts external to our a priori grasp, such as causal or historical facts. From this externalist perspective, there is no reason to suppose that when any two atomic concepts latch onto the same entity this identity will be a priori transparent. For example, suppose that the facts which give my *Cicero* and *Tully* concepts their semantic value are external to the mind, as supposed by an orthodox Kripkean causal theory of reference. Clearly this leaves it open that I could possess both concepts and yet not know that they co-refer.

Reference here will be ‘direct’ in that it does not depend on any descriptive or other complexity in the relevant concepts, but this kind of directness is no guarantee at all that relations of identity and other essential properties of the referents will be transparent to me. So, in the case at hand, both *pain* and *C-fibres* might similarly refer directly to the same entity, yet this not be transparent to someone who possesses these concepts (cf. Levine, 2001, ch. 2; Papineau, 2007, sect. 4).

## 2. Contrary Intuitions

Some of you might be getting impatient. Do we really need an *argument* to show that pains are not the same as C-fibre firings? Isn’t it *obvious* that conscious states are distinct from straightforward physical states?

I agree entirely that this *seems* obvious. There is no doubt that we are all subject to a strong ‘intuition of distinctness’. How can feelings — the colours, the smells, the excitement — possibly be the same as grey mushy brain states?

However, it is philosophically very unwise to rest one’s case on unanalysed intuitions. Intuitions can be quite mistaken, and they need to be examined very carefully before arguments can be grounded on them.

It seems obvious that the earth is stationary. It seems incontrovertible that time moves. And so on. If intuitions like these were allowed to stand, many important theories would be strangled at birth.

Of course argument has to stop somewhere. If every assumption always has to be scrutinized and assessed, then we will be off on an interminable regress. Some claims must be taken as given in any argumentative context. But this does not justify resting a case on an intuition when that intuition runs counter to an otherwise well-supported theory. Given such a theory-intuition clash, the intuition has the status of a disputed premise. So we need to subject it to further assessment, hoping either to show that it follows from further agreed assumptions, or alternatively to explain why it arises even though it is false. (Consider, for example, how Galileo showed why the earth appears to be stationary even though it is moving, and B-series theorists aim to show why we have the impression of a moving present even though there is no such thing.)

I take the intuition of mind-brain distinctness to raise just this kind of issue. On the one hand we have straightforward physicalism, which is strongly backed by causal-explanatory considerations. On the other hand we have the brute intuition that mind and brain must be distinct.

Given this, we need to know whether there is any independent support for this intuition or whether it stems from some fallacious source.

This is why it is necessary to consider whether there are any good *arguments* that might back up the intuition of mind-brain distinctness, as I did in the previous section. My verdict there was that this intuition of distinctness lacks independent argumentative support. Perhaps, then, we should take seriously the possibility that it is ungrounded.

I myself think this is the right diagnosis. In my view, the intuition of distinctness is a confusion engendered by the peculiar way in which we think about consciousness. Consider what happens when we think about some conscious state *as* a conscious state. For example, we think about what it is *like* to feel pain or to see something as red. Such thoughts are characteristically accompanied by some version of the conscious state being thought about. Thus for example, we might think about the experience of seeing something red while we are actually having that experience; alternatively, we might think about what it is like to see red while we are imaginatively recreating the experience. I shall call thoughts of these kinds ‘phenomenal thoughts’.<sup>3</sup> Their characteristic feature is that the conscious *referent* itself is involved in the *vehicle* of thought. We can think of this as the ‘use-mention feature’ of phenomenal thoughts: phenomenal thoughts use the very states that they mention

Now, this use-mention feature carries much potential for confusion. In fact, I think that there are a number of *different* ways in which it can seduce thinkers into an intuition of dualism. Here I want to focus on a line of thought that seems to me to have some affinity with Strawson’s ‘real physicalism’.<sup>4</sup>

When we reflect introspectively on phenomenal thoughts, we become aware that the thought is accompanied by the very experience it refers to, or an imaginative recreation thereof. (The conscious feeling is *right there* in our thinking.) This can lend credibility to the idea that phenomenal thinking gives us a special acquaintance with reality — here at least the object of thought are ‘given’ to us. For, to repeat, in phenomenal thought the conscious referent seems to be present in the thinking itself, without any veil between subject and

[3] Sometimes this kind of thought is said to involve a distinctive species of ‘phenomenal concept’ (cf. Loar, 1990; Papineau, 2002, ch. 2). But this characterization is not essential to the point I wish to make. It is enough that, in a familiar range of cases, thoughts about conscious states are accompanied by versions of those same states, whether or not we view these thoughts as involving special phenomenal concepts.

[4] For other ways in which the use-mention feature of phenomenal thoughts might generate an intuition of distinctness, see Papineau (2002), ch. 6; and Melnyk (2003).

object. So we suppose that the nature of conscious states must be completely transparent to phenomenal thinking — nothing essential will be omitted. From which it follows that conscious states cannot be identical to physical states, at least not to straightforward physical states. For it is certainly not immediately apparent to phenomenal thought that pains are identical to C-fibres firing or any other straightforward physical state.

I think that this whole line of thought is misguided. There is no such thing as unmediated reference, even in the phenomenal realm. Phenomenal thoughts do not have any magical power to reach out and grasp their objects transparently. Phenomenal thoughts, just like any other intentional states, gain their referential powers from causal and historical relations, and those referential relations can leave many essential features of the referents opaque.

Still, even if the idea of epistemologically transparent phenomenal reference is misguided, we can still see how it might seduce us into the intuition that the mind must be distinct from the brain. For conscious introspection does indeed show us that conscious states are present in our phenomenal thinking, in a way that the referents of other thoughts are not; and it is a natural enough step to infer from this that phenomenal thinking, unlike other thoughts, somehow casts a pure light on conscious reality as it is in itself. This latter step may be fallacious, but this needn't stop it being plausible enough to account for the intuition that consciousness, as revealed to us in phenomenal thought, can't be identical with anything physical.

If this diagnosis is right, then I take it that the opposition to straightforward physicalism falls away. On the one hand we have the causal-explanatory arguments weighing strongly in favour of straightforward physicalism. On the other, we have the brute intuition that straightforward physicalism cannot be true. But if this intuition can be explained away, in the sense of showing how it would still arise even if straightforward physicalism were true, then the causal-explanatory arguments stand unopposed. Those who feel that they should still stand by their anti-physicalist intuition, even in the face of this explanation, would do well to consider the fate of other anti-theoretical intuitions, like the intuition that earth stands still.

### **3. Real Physicalism and Panpsychism**

Those who believe that phenomenal thought offers unmediated acquaintance with conscious reality don't necessarily reject all varieties of physicalism. True, they will reject, along with Strawson, any physicalism

that supposes that straightforward terms like *C-fibres* ‘can fully capture the nature or essence of experience’. But that leaves room for a form of physicalism (‘real physicalism’) in which physical states are identified in some superior way which makes their identity with conscious states manifest.

It is not immediately obvious, however, exactly how the position which then results differs from straightforward physicalism. Even though straightforward physicalists do not share the motivations of real physicalists, they are likely to agree with them on the following two propositions.

(1) The terms of current physics, like *C-fibres*, or indeed *quark*, pick out their referents in ways that fail to make all their essential features transparent. That is why claims like *C-fibres* = *pain* cannot be known a priori. Moreover, any terminology developed by future physical theories is likely to be similar in this respect.

(2) An identification of some real physical state with a given conscious experience will only be transparent to the mind if we conceive of that state in phenomenal terms — that is, *as* that conscious experience.

In short, straightforward physicalists are likely to agree with real physicalists that scientific talk of relevant brain states picks out states which are in fact essentially conscious, but does not a priori display those states as conscious.

True, straightforward physicalists will attach far less significance to this epistemological opacity than real physicalists, for straightforward physicalists do not accept *any* transparent modes of thought, and in particular do not accept that phenomenal thought is superior in this respect. So from their perspective there is nothing epistemologically second-class about scientific terminology — for no other mode of thought makes all essential properties transparent either. (That is why I am happy to say that ordinary scientific terms ‘can fully capture the nature or essence of experience’. I don’t recognize any way in which the mind ‘captures’ something, apart from simply referring to it.)

Still, despite this difference of perspective, straightforward physicalists will agree with real physicalists that the metaphysics of mind involves states that are inherently conscious, but that are not displayed as conscious by straightforward physical terminology.

This raises the question of what makes real physicalism a distinctive metaphysical position. The answer, I take it, is to do with panpsychism. Real physicalism, but not straightforward physicalism, implies panpsychism.

Strawson holds that consciousness cannot be constituted out of materials that are not conscious, in the way that liquidity can be

constituted out of materials that are not liquid. So he concludes that complex conscious states, like states of human awareness, must have simple components that are themselves experiential in nature.

From my point of view, this micropsychism (and the panpsychism that Strawson quickly infers from it) is unmotivated, for reasons I will spell out in a moment. But first I would like briefly to point to two internal difficulties that seem to face Strawson's view.

First, it is not clear that Strawson's micropsychism can play the role it is invoked to fill, of explaining how complex conscious states can be constituted out of simpler components. Strawson says:

It is at this point, when we consider the difference between macroexperiential and microexperiential phenomena, that the notion of emergence begins to recover some respectability in its application to the case of experience. (Strawson, 2006, p. 27)

Somehow, the experientiality of the microcomponents is supposed to help explain the emergence of the macroexperiential states. But how exactly will this work? Explanations of normal physical macrophenomena like liquidity hinge on some understanding of the macrophenomenon's *causal role* — we can show how liquidity arises from certain microarrangements by showing how those arrangements ensure that liquidity's causal role is filled. But when we think phenomenally about macroexperiential events like seeing something green, or listening to a high-pitched noise, we do not conceive these macroexperiences in terms of any causal role. This makes it quite obscure exactly how their emergence is to be explained in terms of microcomponents whose simple experiential nature is quite different. Why should certain arrangements of these simple experiential components give rise to the experience of seeing something green? This would seem just as 'brute' as the straightforward physicalist explanations that Strawson is inclined to reject.

Second, the idea that complex conscious states are composed of simpler experiential components would seem to be in tension with the idea that phenomenal thinking gives us unmediated epistemological access to the nature of conscious experience. If the experience of seeing something green is a complex state, composed of experiential simples, then this complexity is presumably an essential property of the macroexperience. Yet phenomenal thinking about the experience of green does not reveal it to be some kind of structured complex, in the way it ought to if it is to lay bare all the essential features of experience.

I said that Strawson's panpsychism is unmotivated. Let me now explain why. Strawson needs panpsychism because he takes there to be an unbridgeable divide between the non-experiential and experiential realms. As he sees it, the whole cloth of experience cannot possibly be woven from threads which are themselves non-experiential. So Strawson is driven to posit that the simples from which complex mental states are built are themselves experiential in nature.

However, this whole line of thought is driven by his initial rejection of straightforward physicalism. To see this, consider how the issue of emergence will appear if we do not abandon straightforward physicalism. According to straightforward physicalism, the general phenomenal category of *being conscious*, like more specific phenomenal categories, will be identified with some broadly physical category (such as activity in a global workspace, or representation that controls action, or oscillations at 40 hertz in the sensory cortex . . .). Now, presumably there will be no difficulty in explaining how such a complex physical category can be constituted out of non-experiential physical simples. This would be just the kind of explanation that Strawson allows can account for the emergence of liquidity from non-liquidity. But this will now amount to an explanation of consciousness in terms of non-experiential physical simples. For if consciousness is known to be identical to some complex physical category, and we can explain this category in terms of non-experiential simples, then we will therewith have explained consciousness.

It might seem as if straightforward physicalism still owes some further explanation that can't be given in straightforwardly physical terms. *Why* is it like something to have a brain state that is active in a global workspace, say? Why are just those brain states conscious, and others not? Surely this is just the same mystery as motivates Strawson; and surely it cannot be answered simply by showing how global workspace activity can be realized by some complicated arrangement of non-experiential physical simples.

However, the demand for further explanation depends on the intuition that conscious mind and physical brain are distinct. If we did not have this intuition, we would not feel there was anything more to explain. Perhaps the point is easiest to see with some specific phenomenal category, like pain, say. Suppose, for the sake of the argument, that pain is identified with C-fibre firings. Do we still need to explain why C-fibre firings 'give rise to' pain? I say not. If C-fibre firings *are* pains, then there is no remaining question of *why* pains are found where C-fibres are firing. Given that's what C-fibre firings are, there's

no possibility of their being otherwise. To ask for further explanation illegitimately presupposes that the pain is distinct from the brain state.

Sometimes people accept this point for pains and other specific phenomenal categories but resist it for the general category of consciousness as such. Why should it be just *that* range of physical states, and not some other, that ‘gives rise to’ feelings? But the straightforward physicalist should simply say the same thing again. If activity in a global workspace *is* consciousness, there’s no issue of explaining *why* it is — it couldn’t have been otherwise. The request for explanation arises only as long as we remain in the grip of the intuition that being conscious must be extra to any straightforward physical property.

So I see nothing wrong with explaining the ‘emergence’ of consciousness from non-experiential physical simples. If the property of being conscious is identical to some straightforward physical property, there can be no barrier to such an explanation. We are only driven towards panpsychism if we posit a radical divide between the experiential and non-experiential realms. Straightforward physicalism rejects any such divide. Those who listen to argument, and ignore brute intuition, can thus steer clear of panpsychism.

### References

- Chalmers, D. (1996), *The Conscious Mind* (Oxford: Oxford University Press).  
 Jackson, F. (1998), *From Metaphysics to Ethics* (Oxford: Oxford University Press).  
 Levine, J. (2001), *Purple Haze* (Oxford: Oxford University Press).  
 Loar, B. (1990), ‘Phenomenal states’, in *Philosophical Perspectives*, 4, ed. J. Tomberlin.  
 Melnyk, A. (2003), ‘Papineau on the intuition of distinctness’, *SWIF Forum on Thinking about Consciousness*, [http://lgxserver.uniba.it/lei/mind/forums/004\\_0003.htm](http://lgxserver.uniba.it/lei/mind/forums/004_0003.htm)  
 Papineau, D. (2002), *Thinking about Consciousness* (Oxford: Oxford University Press).  
 Papineau, D. (2006), ‘Phenomenal and perceptual concepts’, in *Phenomenal Concepts*, ed. T. Alter and S. Walter (Oxford: Oxford University Press).  
 Strawson, G. (1994), *Mental Reality* (Cambridge, MA: MIT Press).  
 Strawson, G. (2006), ‘Realistic monism: Why physicalism entails panpsychism’, *Journal of Consciousness Studies*, 13 (10–11), pp. 3–31.