

Peter Carruthers
and Elizabeth Schechter¹

Can Panpsychism Bridge the Explanatory Gap?

Strawson (2006) claims that so long as we take the physical ultimates of the world to be non-experiential in nature, we will never be capable of explaining or understanding how human conscious experience can emerge from physical processes. He therefore urges physicalists to embrace a panpsychist metaphysics. This is because, he says, we can better understand how macro-experientiality (the conscious experience of creatures like us) might arise from micro-experientiality (the conscious experientiality of the physical ‘ultimates’ of the universe) than we could ever understand how experientiality arises from the non-experiential. Strawson’s view is that the experiential features of the ultimates can make it intelligible that there should be a scientific explanation or reduction of human consciousness. For he says that once we have accepted panpsychism, ‘the notion of emergence begins to recover some respectability in its application to the case of experience’ (p. 27),² and that he is proposing a ‘general framework of thought in which there need be no more sense of a radically unintelligible transition in the case of experientiality than there is in the case of liquidity’ (p. 28).

The panpsychist metaphysics that Strawson proposes is intended, therefore, to help us with the mind/body problem: the problem of how human conscious experience relates to the physical matter of the brain. He says that many physicalists take the claim that the experiential is also physical as nevertheless:

profoundly problematic *given what we know about the nature of the physical*. But they have already made a large and fatal mistake. This is

[1] The ordering of the authors’ names is alphabetical.

[2] All page references are to Strawson (2006).

because we have no good reason to think that we know anything about the physical that gives us any reason to find any problem in the idea that experiential phenomena are physical phenomena (p. 4, emphasis in original).

In fact, Strawson blames physicalists for creating an insoluble problem (insoluble within the physicalist's current metaphysical framework) where there really is none, via their dualistic tendency to set up a dichotomy between the 'mental' (i.e. conscious, experiential) and the 'physical'. This is a false dichotomy, Strawson says: it is all physical, and we have no reason to think that it isn't all experiential as well. (Physics itself is silent on this matter, he believes.) If we only accepted that the physical was in and of itself essentially experiential, then there would be no mind/body problem *per se*, and scientific explanations of human consciousness (of macro-experientiality) would become intelligible.

One major manifestation of the mind/body problem is the explanatory gap that, it is claimed, will always exist between any description of the physical and/or functional properties of the human body and brain, on the one hand, and consciousness described in phenomenal or experiential terms, on the other. Those who take the explanatory gap to be a serious problem do so precisely because the gap doesn't seem to exist between other higher-level properties (non-consciousness involving properties) and the phenomena that *they* reduce to. Strawson uses the example of liquidity to illustrate the intelligibility of most purported cases of emergence, and to provide a contrast between such cases and the case of consciousness. Intelligible scientific explanations and reductions of such phenomena provide us with a feeling of necessity: given the various chemical properties of H₂O molecules, and given that water, a liquid, is composed of H₂O molecules, and given that liquids are forms of matter in which the molecules that compose them slide off each other instead of gripping each other tightly in a lattice, we see that the property of liquidity *must* emerge from the properties of H₂O molecules, even though those molecules aren't themselves liquids. We see that water's property of liquidity is entailed by the properties of the (non-liquid) H₂O molecules that compose it. But no non-experiential properties seem to compel or entail the qualitative aspects of any particular conscious experience. Whatever the non-experiential features of my body and brain are, it will always appear to me that the experiential features of my consciousness could still have been different, or even absent altogether.

Anti-physicalists like Chalmers (1996) often take this explanatory gap as demonstrating that there is a *metaphysical* gap between the

experiential and the (non-experiential) physical. Such philosophers — who say that the mind/body problem is insoluble precisely because experiential properties really *aren't* composed of non-experiential ones — are often called 'property dualists'. Strawson is eager to distance himself from this kind of anti-physicalist, however, and is at pains to point out that he believes that all properties, including experiential properties, are physical. But he thinks that if the ultimate entities in the universe are, while physical, at the same time experiential, then the explanatory gap can be closed, in principle.

One of us (Carruthers, 2000, 2005) has previously suggested, however — along with Loar (1990), Tye (2000) and others — that the existence of the explanatory gap is best explained not by giving up on standard forms of physicalism, but by focusing on the distinction between phenomenal *concepts* and other kinds of concept. Most concepts are, in a broad sense, functional, defined by their relationships with other concepts and/or embedded in bodies of belief about the world that modulate their use. In contrast, recognitional concepts, like 'red', can be applied directly on the basis of perceptual or quasi-perceptual acquaintance with their instances. And phenomenal concepts like 'feels itchy' or 'seems red' are *purely* recognitional concepts. In contrast to non-purely-recognitional concepts like 'red', concepts like 'seems red' aren't tied to any other concepts or beliefs about the world. While we wouldn't apply the concept 'red' to a wall that looked red, for example, if we knew that we were standing in a room illuminated by a red light, our purely recognitional concept 'seems red' would still apply. Regardless of what one knew about the lighting conditions in the room — and in fact regardless of what one knew about *anything* in the world or our own bodies — the concept 'seems red' would still apply solely on the basis of our phenomenal experience of the wall.

When we think about the qualitative nature of our experience we use phenomenal concepts, and these concepts are always isolable from whatever physical concepts might be employed in any proposed physicalist explanation of consciousness (whether neurological, functional or representational). Our phenomenal descriptions of experience — laden, as they are, with phenomenal concepts — can therefore never be entailed by descriptions of our experience that don't use those very same concepts;³ our phenomenal concepts are free to drift, as it were, unanchored. Accordingly it will always be possible to feel

[3] There exist some *a priori* relations amongst phenomenal concepts that permit a weak sort of explanation-by-constitution; but these are extremely limited. For example, if I undergo a feels-red experience then I know *a priori* that I am undergoing a feels-visual experience.

that phenomenal consciousness *itself* isn't anchored by, and that it doesn't depend upon, any other features of this world. But that is just because we have ways of thinking about our experience that don't depend on the application of any non-recognitional concepts. The explanatory gap is here revealed as 'a cognitive illusion' (Tye, 2000), resulting from the conceptual isolation of the concepts that we employ when thinking about our own experience.

Strawson makes no mention of this sort of indirect strategy for dissolving the explanatory gap. (The strategy is indirect because it doesn't seek to explain why such-and-such a physical system should feel like *this*; rather, it seeks to explain why no such explanation is forthcoming, in terms that will make the continued demand for an explanation evaporate.) And he is committed to claiming that the mind/body problem can't be solved within the framework provided by standard forms of physicalism. While disagreeing, we propose to set our reservations aside in the discussion that follows. Our purpose here is rather to evaluate whether Strawson's panpsychism leaves us any better off in the search for a direct, full-frontal solution to the mind/body problem. We will argue that it does not.

Before we can embark on that discussion, however, we need to get a sense of what it would mean to accept that the physical ultimates of this world are also experiential in character. If one takes seriously Block's (1995) distinction between access consciousness (consisting of functionally defined states that can enter into a subject's reasoning, have an impact on belief, and/or be reportable in speech, etc.) and phenomenal consciousness (states that have a feel), then there are two conceptually distinct ways for a phenomenon to be experiential. An access-conscious state would be experiential in the sense of being somebody's (or something's) experience, by virtue of playing a special functional role for some conscious being. A phenomenally conscious state, in contrast, would be experiential in the sense of possessing a feel. A feel that isn't felt by any conscious subject then becomes a conceptual possibility, at least. Indeed some philosophers believe that our experiences possess intrinsic, irreducible, non-representational, non-relationally-defined properties, called *qualia*. Many philosophers argue that qualia are metaphysically impossible, or, at least, that they don't exist in this universe. But those who believe in the existence of qualia assert that the universe contains not only

Accordingly, I can explain how it is that I am undergoing a feels-visual experience by saying that I am undergoing a feels-red one. It should be plain, however, that very little about my experience can be explained in this way, i.e. even by using other phenomenal concepts.

people who feel pain and felt pains, but also pain-feely properties, pain qualia, themselves.

Strawson's proposed panpsychism is one according to which the ultimates (quarks or strings or whatever) are tiny subjects of experience, for he says that 'experience is impossible without an experiencer', and that panpsychism means that we have 'a rather large number of subjects of experience on our hands' (p. 26). However, he might instead have proposed a metaphysics according to which the physical ultimates of this world aren't conscious subjects, but rather have feel-properties attached to them. On this version of panpsychism, the ultimates are experiential entities in the sense that they possess irreducible properties of experience, or qualia, but are not themselves subjects of experience.

We believe that the version of panpsychism Strawson advocates, according to which the ultimates are themselves subjects of experience, is the more extreme and more problematic version, and lays itself open to a greater number of objections. We will proceed, then, by evaluating the weaker and more plausible version, according to which the ultimates of the world possess qualia. Any objections that we present to this weaker version will apply also to the stronger version (to which additional objections apply as well).⁴

Can macro-experientiality be reductively explained in terms of micro-experientiality, then? It is hard to see how it can be. How could trillions of particles, whatever their experiential nature, constitute what feels like (and what Strawson [1997] has argued we have phenomenological warrant to believe *is*) a single subject of experience? One problem here is epistemological: I can't know anything about the experientiality of the ultimates that compose me. Certainly I can't know the experiential properties of the individual atoms and molecules that constitute my brain on the basis of introspection. And it is hard to see how one could, even in principle, get any evidence as to the nature of those properties. In which case we can see in advance that we could never be in a position to mount a reductive explanation of our experience. We shall never be able to start from the known experiential properties of the ultimates that compose us, deducing that our experience should have the character that it does, in the way that we *can* start from the known properties of H₂O molecules and deduce

[4] As Strawson himself admits, we would 'need to address William James's well known objection to the idea that many subjects of experience can somehow constitute a single "larger" subject of experience'. Although Strawson protests that there is 'no more difficulty in the idea that the experiential quality of micro-experientiality is unimaginable by us than there is in the idea that there may be sensory modalities (qualitatively) unimaginable by us' (p. 27), at least we know that bats, for example, *have* sensory modalities.

the liquidity of water. If we can't know the experiential properties of the ultimates, then we can't ever provide a reductive explanation of phenomenal consciousness in terms of such properties, either.

Strawson will probably reply that he is concerned with a metaphysical rather than an epistemic sense of 'explain'. At issue is not whether we can know, or have reason to believe in, some particular reductive explanation of the properties of our experience. It is rather whether there *exists* such an explanation, in the world. Strawson might say that we can see in advance that there can't (on metaphysical grounds) be any explanation of phenomenal consciousness in terms of standard physical properties like mass, electric charge and so forth. For these properties and the properties of our experience are too heterogeneous to admit of explanation of the latter by the former. Whereas if panpsychism is adopted, and the physical ultimates are at the same time experiential, then at least we shall be explaining like with like. Even if we don't, and can't, know enough about the experiential properties of the ultimates to construct a detailed explanation, we can at least see that one isn't ruled out in principle by the metaphysics of the situation.

Strawson might also reply (either instead, or in addition) that the experiential properties of the ultimates might in principle be known on the basis of an inference to the best explanation, where the target of explanation is the character of our own conscious experience. If by postulating that the atoms in my brain possess such-and-such qualia, and by adding that those atoms with those properties are interacting thus-and-so, we can explain why my experience should feel to me the way it does, then this would provide us with good reason to believe that the atoms in my brain do indeed possess those properties. This sort of indirect inference to properties that can't be accessed directly is just an application of standard abductive inference in science.

While we have doubts about each of these two lines of response to our challenge, we shan't press them here. Rather, we will show that even in a best-case scenario — in which the phenomenal properties of the ultimates are known in complete detail — panpsychism still wouldn't help us with the mind/body problem. For the explanatory gap would remain in place, untouched and wide open as ever: panpsychism does nothing to close it.

To begin to see this, just ask yourself whether, if everyone in the world were a committed panpsychist, children would cease wondering whether their best friends shared their phenomenal colour experiences when visually confronted with the same scenes. Or ask yourself if Tye and Chalmers would co-author papers on the impossibility of zombies. Why would they? If zombies are conceivable now, then they

will remain conceivable no matter what they, and we, are made of. It will remain conceivable that there should be a zombie made out of *exactly the same stuff as I am* (whatever that stuff is like), but without undergoing an experience like *this* one (like the one I am currently having). In fact it will always be possible for the panpsychist to think thoughts like the following:

Couldn't my brain be in the exact (non-experientially described) state that it is in now, as I look at this tree, and couldn't the particles that compose me possess the very same (*experientially* described) properties that they have, and yet couldn't *I* be in a *different* experiential state? Even if every ultimate particle in my body possessed a green-feely experiential property, couldn't *I* be undergoing an experience with the phenomenal properties of an experience of red? And likewise, couldn't there be someone who was composed of particles exactly like mine arranged in exactly the same way, each of which possessed the very same qualia as do the particles that compose me, yet who lacked phenomenal consciousness altogether (just as, according to Strawson's proposed metaphysics, a stone is composed of the exact same experiential ultimates yet lacks experientiality or consciousness of its own)?

The point is a simple one, and was made long ago by Block (1978) in his population-of-China counter-example to functionalism. If one imagines millions of components interacting together in ways that mirror the interactions of my own components, it is still possible to conceive that the resulting complex being lacks phenomenal consciousness altogether, even if each of those components possesses phenomenal properties. In which case no amount of knowledge of the feely nature of the ultimate physical particles could ever explain phenomenal consciousness in the way that knowledge of the component particles of water and the manner in which they interact can explain liquidity.

There are familiar avenues of reply to the population-of-China counter-example, of course. For example, Strawson might retort that the problem is merely one of the limits of our imagination. Since we cannot really conceive, in detail, of millions and millions of experiential entities interacting in as-yet-to-be-specified and highly complex ways, we cannot really tell whether some such story mightn't constitute a successful reductive explanation of our own experientiality. But this sort of reply is equally available to defenders of standard (non-experiential) forms of physicalism, of course (Dennett, 1991). So if it works, it would only serve to undermine Strawson's own argument for panpsychism.

But isn't there at least *some* explanatory gain to be had from panpsychism? Doesn't the fact that the phenomenal properties attributed

to the ultimate particles of the universe are of the same metaphysical *type* as the target to be explained (our own phenomenal consciousness) at least indicate that the latter might be explicable *in principle*? In fact not. For we have not the faintest idea how the phenomenal properties possessed by one entity or set of entities might contribute to a reductive explanation of the phenomenal properties possessed by another, as the thought experiments described above indicate.

Panpsychism was urged on us as the one move that might enable us to see how the explanatory gap could one day be bridged. But it turns out to be a blind alley. Even if the ultimates of the universe are experiential in nature, the explanatory gap remains untouched. It is better, then, to remain an old-fashioned (non-panpsychic) physicalist, and to accommodate or circumvent the explanatory gap by other means. Indeed, it is worth noting that the sort of indirect approach to explaining the explanatory gap that we mentioned earlier also has the resources to explain why a panpsychist metaphysic leaves the explanatory gap intact. The claim is that phenomenal concepts (the concepts by means of which we think about phenomenal properties) are conceptually isolated ones, lacking *a priori* connections with other concepts, and lacking any embedding in a wider theory. In which case no story — no matter how detailed — about the component parts of a creature and their modes of interaction will entail an application of such a concept, even, as it turns out, if the story utilizes phenomenal concepts themselves to ascribe feely properties to the components.

References

- Block, N. (1978), 'Troubles with functionalism', in *Minnesota Studies in the Philosophy of Science*, ed. C. Savage, **9**, pp. 261–325.
- Block, N. (1995), 'A confusion about the function of consciousness', *Behavioral and Brain Sciences*, **18**, pp. 227–47.
- Carruthers, P. (2000), *Phenomenal Consciousness* (Cambridge: Cambridge University Press).
- Carruthers, P. (2005), *Consciousness: Essays from a Higher-Order Perspective* (Oxford: Oxford University Press).
- Chalmers, D. (1996), *The Conscious Mind* (New York: Oxford University Press).
- Dennett, D. (1991), *Consciousness Explained* (Allen Lane).
- Loar, B. (1990), 'Phenomenal states', in *Philosophical Perspectives: Action Theory and Philosophy of Mind*, ed. J. Tomberlin (Ridgeview).
- Strawson, G. (1997), 'The self', *Journal of Consciousness Studies*, **4** (5–6), pp. 405–28.
- Strawson, G. (2006), 'Realistic monism: Why physicalism entails panpsychism', *Journal of Consciousness Studies*, **13** (10–11), pp. 3–31.
- Tye, M. (2000), *Consciousness, Color, and Content* (MIT Press).

to the ultimate particles of the universe are of the same metaphysical *type* as the target to be explained (our own phenomenal consciousness) at least indicate that the latter might be explicable *in principle*? In fact not. For we have not the faintest idea how the phenomenal properties possessed by one entity or set of entities might contribute to a reductive explanation of the phenomenal properties possessed by another, as the thought experiments described above indicate.

Panpsychism was urged on us as the one move that might enable us to see how the explanatory gap could one day be bridged. But it turns out to be a blind alley. Even if the ultimates of the universe are experiential in nature, the explanatory gap remains untouched. It is better, then, to remain an old-fashioned (non-panpsychic) physicalist, and to accommodate or circumvent the explanatory gap by other means. Indeed, it is worth noting that the sort of indirect approach to explaining the explanatory gap that we mentioned earlier also has the resources to explain why a panpsychist metaphysic leaves the explanatory gap intact. The claim is that phenomenal concepts (the concepts by means of which we think about phenomenal properties) are conceptually isolated ones, lacking *a priori* connections with other concepts, and lacking any embedding in a wider theory. In which case no story — no matter how detailed — about the component parts of a creature and their modes of interaction will entail an application of such a concept, even, as it turns out, if the story utilizes phenomenal concepts themselves to ascribe feely properties to the components.

References

- Block, N. (1978), 'Troubles with functionalism', in *Minnesota Studies in the Philosophy of Science*, ed. C. Savage, **9**, pp. 261–325.
- Block, N. (1995), 'A confusion about the function of consciousness', *Behavioral and Brain Sciences*, **18**, pp. 227–47.
- Carruthers, P. (2000), *Phenomenal Consciousness* (Cambridge: Cambridge University Press).
- Carruthers, P. (2005), *Consciousness: Essays from a Higher-Order Perspective* (Oxford: Oxford University Press).
- Chalmers, D. (1996), *The Conscious Mind* (New York: Oxford University Press).
- Dennett, D. (1991), *Consciousness Explained* (Allen Lane).
- Loar, B. (1990), 'Phenomenal states', in *Philosophical Perspectives: Action Theory and Philosophy of Mind*, ed. J. Tomberlin (Ridgeview).
- Strawson, G. (1997), 'The self', *Journal of Consciousness Studies*, **4** (5–6), pp. 405–28.
- Strawson, G. (2006), 'Realistic monism: Why physicalism entails panpsychism', *Journal of Consciousness Studies*, **13** (10–11), pp. 3–31.
- Tye, M. (2000), *Consciousness, Color, and Content* (MIT Press).