21

CAN WE SUM SUBJECTS? EVALUATING PANPSYCHISM'S HARD PROBLEM

Luke Roelofs

Panpsychist accounts of consciousness hold that humans are conscious because matter itself is conscious. One widespread motivation for panpsychism is dissatisfaction with physicalist explanations of consciousness (Nagel 1979; Seager 1995; Chalmers 1995; Strawson 2006). If a non-mental understanding of physical matter leaves an 'explanatory gap' between physics and consciousness, and all physicalistic attempts to 'close the gap' fail, perhaps we should make consciousness itself fundamental, governed by some set of psychophysical laws of nature. Since the laws of physics seem to derive much of their appeal from being simple and general, with a multitude of forms building up gradually from a small set of widespread basic elements, let us suppose similarly general psychophysical laws, with consciousness throughout the material universe.

But what if we accept fundamental, omnipresent, consciousness but find our explanatory situation unchanged: human consciousness stills seems like an unexplained mystery? This is the impasse to which many critics have accused panpsychism of leading. Since panpsychism aims to add new fundamental laws *only* at the basic level, it still needs to explain how complex things like human minds are 'built up' from the basic experientiality of matter. And it needs to offer a more satisfying explanation than either physicalism, which denies the fundamentality of consciousness, or emergentism, which denies its omnipresence.[1] And there is a concern, widely voiced by both opponents of and sympathisers with panpsychism,[2] that this cannot be done: minds simply do not combine in the necessary way. This has been labelled the 'combination problem' (Seager 1995: 283).

Different panpsychists respond differently to the combination problem, and different components of it demand different responses. This chapter focuses on the very possibility of explanatory relations between distinct subjects; in section 1 I explain why this is the principal 'hard problem' of combination. Section 2 covers the standards for an adequate solution, section 3 reviews three proposals, evaluating them in light of these standards, and section 4 cautiously recommends a mixture of two of these proposals.

1. Hard and Easy Problems of Combination

Much work has been done to classify the components of the combination problem. Coleman distinguishes 'internal' and 'bridging' problems (2017: 3 ff.), Goff distinguishes 'from above' and 'from below' (2017b: chs. 7–8), and Chalmers distinguishes problems around subjecthood, qualities, and structure (2017). I will suggest another division, modelled on Chalmers' division between the 'hard

problem' and 'easy problems' of consciousness (1995: 1 ff.). I believe there is a 'hard problem of combination', interestingly different from several 'easy problems of combination'.

Neither Chalmers' distinction nor mine is about *degree* of difficulty, but rather about *type* of difficulty. Chalmers says: 'The easy problems of consciousness are those that seem directly susceptible to the standard methods of cognitive science. . . . The hard problems are those that seem to resist those methods' (1995: 201).

Even if it is very difficult to say, for instance, how the brain produces words matching the stimuli its sense organs are exposed to, we at least have methods for approaching the task: those of psychology, neuroscience, and other empirical sciences of the mind. By contrast, the problem of subjective experience is 'hard' precisely in that those methods seem inappropriate. They just don't make contact with the problem, since for any neural mechanism or functional architecture, it still makes sense to ask 'But why does that *feel like anything*?'

My distinction between 'hard' and 'easy' problems of combination is meant to exhibit an analogous contrast between difficulties in applying an appropriate method, and difficulties where we seem to lack a method which would even address the issue. But the methods in question are different – the easy problems of consciousness are those which are tractable by the methods of the cognitive sciences, but the easy problems of combination are those which are tractable by the methods of *phenomenological analysis*. The hard problems of combination are whatever problems are not thus tractable: I believe the main, and perhaps only, example is what has become known as 'the subject-summing problem'.

By 'phenomenological analysis' I simply mean the inward-looking sort of attention to one's own experience which we routinely use to address questions about whether one type of experience is necessary or sufficient for another. When we ask, for instance, whether the experience of mathematical reasoning is fully accounted for by the experience of entertaining the associated images, or whether an experience of something as objectively existing might require certain kinds of imaginative capacities, we are exploring relations among types of experience by analysing them as they occur within our own mind. While some question the trustworthiness of such methods (e.g. Dennett 1991; Blackmore 2002), they are a well-established part of how philosophers approach questions about experience, not an *ad hoc* invention of panpsychists.

So what are 'hard' and 'easy' problems of combination? Consider the subject-summing problem. This problem is very simple to express: no facts about one subject's consciousness seem to directly explain facts about another's. Or at least, it is not clear how they would, and many people find it implausible that they ever could – in particular, many feel that subjects are in some sense 'metaphysically insulated', 'cut off' from each other (e.g. James 1890: 226; Coleman 2012: 146; Blamauer this volume). Since my parts are not identical with me, their consciousness is cut off from explaining mine. Note that this worry is immune to any resolution based simply in phenomenological analysis of experience, for in phenomenological analysis the experiences I analyse *are all mine*: I cannot do phenomenological analysis on your experience. All I can find is that when certain experiences are had together *by the same subject*, another sort of experience is had *by that subject* – which is irrelevant to explanatory relations among different subjects. Just like the hard problem of consciousness, the subject-summing problem is methodologically distinctive in a way that makes it obscure how to even approach it.

To see the contrast with other problems of combination, imagine we fully satisfied ourselves that given a certain configuration of microsubjects, the whole they compose must be a subject of experience. Many problems remain, but all concern explaining particular kinds of experience – why is the macrosubject's experience unified, qualitatively diverse, coarse-grained, epistemically bounded, and so on? And for these questions, phenomenological analysis is entirely relevant, since phenomenological analysis is how we decide what sorts of experience might give rise to what other sorts. For example, for the 'palette problem', of how a few basic ingredients can yield the diverse qualities we experience, it is relevant to analyse the relations among different experienced qualities, to

ask whether some are blends of others, and so on (see Mizrahi 2009). So for addressing these other problems, panpsychists can draw on and extend existing phenomenological work – work which on each occasion is done within a single subject's mind. This is the sense in which I call them 'easy'.

There might be other hard problems of combination, apart from that of subject-summing. But that is my primary example, and it is to this hard problem that I devote the rest of this chapter.

2. Explanation and Explanatory Gaps

To solve the subject-summing problem, microsubjects must 'explain' macrosubjects. But what is it for one thing (x) to 'explain' another (y)? One answer is that an idealised reasoner, given complete knowledge of x, could deduce everything about y. Moreover, they need no additional *a posteriori* premises: if some additional *a posteriori* premise z is needed to deduce y, then what explains y is 'x and z together', not just x. Call this the '*a priori* deduction' standard of explanation. It follows that x fails to explain y if 'x-without-y' is conceivable, for deduction just rules out what cannot be consistently conceived. Hence anti-physicalists argue for an explanatory gap from the conceivability of any physical facts without consciousness.

Applying the same standard, we ask: given a complete account of the microsubjects, is it conceivable that there be no consciousness in the whole? Note that 'a complete account of the microsubjects' covers not only their experiential properties, but also the relations among them, their physical properties, and the micro-level laws that govern them. This includes psychic or psychophysical laws, *if they operate at the micro-level*, but not if they are 'cross-level' laws, connecting facts at one mereological level with facts at another (as 'emergence laws' do).

A priori deduction is not a universally accepted standard for explanation. If we have compelling empirical and theoretical reason to identify A with B, why should it matter if our concept of B gives us no grasp of A? Plausibly, the right model of explanation is whichever best accounts for the explanatory power of natural science (see esp. Block and Stalnaker 1999; Chalmers and Jackson 2001; Diaz-Leon 2011; McQueen 2015), and there can be reasonable disagreement here. We should thus ask whether panpsychists could retreat from the demanding standard of *a priori* deduction, and maintain that microsubjects explain macrosubjects in some other way. But we must recall the panpsychist's dialectical positioning: if physicalism and emergentism are rejected for their explanatory failures, panpsychism needs to deliver more.[3]

So if panpsychists say that microsubjects ground macrosubjects in an 'opaque' but necessary way, without giving any *a priori* insight, then physicalists will reply that their purely physical explanatory base can 'explain' consciousness in that sense (cf. Block and Stalnaker 1999; Levine 1983; Loar 1990). On the other hand, if panpsychists accept the possibility of microsubjects without macrosubjects, and postulate a cross-level emergence law to produce the latter out from former, then emergentists will reply that since we need the law to generate macrosubjects, the microsubjects are pointless.

Are there ways for panpsychists to accept explanatory opacity, or cross-level laws, while retaining an explanatory advantage over physicalism and emergentism? Perhaps. Consider physicalism first. Classic anti-physicalist arguments, such as the 'absent qualia' and 'inverted qualia' thought experiments, are often treated as establishing the same result (the physical does not fix the experiential), but we can actually distinguish two issues. To have no explanatory gap at all, the physical facts need to explain why certain systems are *conscious at all*, and the 'absent qualia' thought experiment seems to undermine this. But they also need to explain why any conscious system has the *particular sort of experiences* that it has, and the 'inverted qualia' thought experiment seems to undermine this too. To explain consciousness means both to 'explain-that' some systems have consciousness, and also to 'explain-what' sorts of consciousness they have.

Explaining-what and explaining-that might come apart. Given x we might be able to deduce that *some* sort of y must appear, and yet not know which specific y. Those who accept the conceivability

of inverted qualia but not absent qualia hold this view of the physical and experiential facts (cf. Shoemaker 1975). On the other hand, *x* might show clearly what sort of *y* there would be *if* there was some sort but leave open that there be none. For instance, suppose we encounter a robotic system of unknown internal structure that seems to converse with us intelligently. Its behaviour might not tell us whether to regard its apparent propositional attitudes as genuine or illusory, but still tell us that *if* it has genuine propositional attitudes, then it has those which it seems to express.

Physicalism seems to neither explain-that nor explain-what, so any view which achieves one or the other would be an explanatory advance. This sense of complete mystery is evoked by Huxley's remark that '[H]ow it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the Djinn, when Aladdin rubbed his lamp' (1986: 193). If the panpsychist can either explain-that macrosubjects have experience, or explain-what sort of experience they have, they will have made human consciousness *less* mysterious than Aladdin's Djinn.

What about maintaining an advantage over emergentism? There are ways to claim partial explanatoriness here as well. One might insist that it is just more intelligible for things of one fundamental type to emerge from each other, than for fundamentally different things to emerge from each other (e.g. Strawson 2006: 16–19), though this may seem question-begging to emergentist opponents. A more neutral way is to focus on theoretical virtues. Why is (non-panpsychist) emergentism unattractive? Not simply because it holds that some of the fundamental laws of nature connect things at different mereological levels, but because doing so compromises its theoretical virtue. This involves three specific problems.

One problem concerns causal exclusion and arises somewhat as follows. If the panpsychist, or the physicalist, could claim that macrosubjects are intelligibly grounded in micro-level entities (identical to them, composed of them, realised by them, etc.) they could maintain that macro- and micro-entities are *not in causal competition*: both can be causally efficacious without their effects being 'overdetermined' like the death of someone shot by two members of a firing squad. But if the micro-level entities need the aid of a special emergence law to generate macrosubjects, it seems like macrosubjects are something ontologically independent of them, and so if both cause some effect that really does look like overdetermination. Insofar as rampant overdetermination is theoretically vicious, this drives emergentists towards either epiphenomenalism about macrosubjects, or denying the causal closure of microphysics. Neither of these is attractive: one attraction of panpsychism is that it seemed to offer anti-physicalists an escape from this dilemma, but re-introducing (strong) emergence will remove this advantage.

Of course, avoiding causal competition between micro and macro is little consolation if there is still competition between micro-experience and microphysics. To avoid this, many panpsychists present their view as a form of 'Russellian monism' (traced to Russell 1927, cf. Eddington 1929; Lockwood 1981; Alter and Nagasawa 2015), on which micro-experience provides the basis for microphysics. On this view, physical descriptions are in some sense 'structural' (cf. Stoljar 2014), saying how matter behaves but leaving unspecified what it actually is that behaves that way. This allows the panpsychist to posit micro-experience as that unspecified inner nature, thereby giving experience a central role in physical causation and avoiding causal competition. But this commits them to a thorough-going isomorphism of mental and physical (relating to the third criticism of emergentism, below), which they cannot violate without risking the loss of their apparent advantage on the causal score.

A second problem is that non-panpsychist emergence laws may offend very heavily against the ideal of 'simplicity'. To preserve the intuition that only human beings and some animals are conscious, the fundamental laws must connect consciousness with a specification – in wholly fundamental terms – of what it takes to be one of the relevant sorts of animal, and this specification would be many orders of magnitude longer than any of the mathematical terms appearing in the laws currently recognised by physics, with many more 'brute facts' about which precise values the many variables must take (Cf. Feigl 1958: 428; Smart 1959: 142–3; Rosenberg 2004: 107–10).

A third issue is the desire to keep the mental and the physical 'in line' with one another as much as possible. Overwhelming empirical evidence seems to show that they are closely intertwined, and even if a theory recognises them as fundamentally distinct aspects, it should not let them 'drift out of sync' (Cf. Mørch 2014: 10, 50). Consequently if the brain's physical profile is just what we would predict, given its parts and their arrangement, we should prefer a theory with a similar sort of continuity in the mental realm. Certainly we should be reluctant to accept a radical change on the mental side if the relevant physical processes are all fully continuous (Cf. James 1890: 147–8; Mørch 2014: 161–3).

If the problem with emergentism is these three threats (to macroexperiential causal efficacy, to theoretical simplicity, and to mental-physical correspondence), then a theory that postulated cross-level laws but avoided these problems would be a comparative explanatory success. And if this theory's microsubjects were crucial to avoiding these results, we would have good reason to prefer panpsychist emergentism to non-panpsychist emergentism.

So if an account of subject combination secures *a priori* entailment of macrosubjects without cross-level laws, it meets the 'gold standard'. Other accounts can be partially successful by doing two things: *explaining more than physicalism* (e.g. by either explaining-that or explaining-what), while being *more theoretically virtuous than emergentism* (in causal non-competition, structural simplicity, and mental-physical correspondence).

3. How to Sum Subjects

This section reviews three approaches to the subject-summing problem that have appeared:

- The 'experience-sharing' approach, on which token experiences belong simultaneously both to parts and whole.
- The 'fusion' approach, in which microsubjects predate macrosubjects rather than co-existing with them.
- The 'phenomenal bonding' approach, on which the crucial explanatory ingredient is a special relation among subjects.

Each approach has strengths and weaknesses: experience-sharing secures theoretical virtue at the expense of revising pre-theoretical notions of subjecthood, fusion sacrifices theoretical virtue almost as much as strong emergentism, and phenomenal bonding rests its explanation on a posit of which we have no positive conception.

Two approaches that I will not discuss here abandon the assumption that elementary particles are the fundamental conscious subjects: 'panprotopsychism' holds that they are fundamental, but not subjects (see Coleman 2012, 2014, cf. Chadha this volume), while 'cosmopsychism' holds that they are not fundamental, but merely the smallest fragments of what is fundamental, namely the conscious cosmos itself (see Jaskolla and Buck 2012; Shani 2015; Goff this volume, cf. Albahari this volume). Rather than trying to solve the subject-summing problem, these approaches seek to re-frame it by changing the starting point. In my view, both struggle to get beyond 'moving the bump under the rug', with the basic difficulty of explaining human-sized subjects always reappearing in a different guise, and with the same basic moves available to address it.

3.1. Experience-Sharing

Why are physical wholes so readily explicable through their parts? One plausible answer is that they have no properties (whether universal types or particular instances) that don't, in some fashion, 'come from' their parts. For example, a physical whole with a red spot shares an instance of redness with a

part of its surface; if dented or ripped, it likewise shares an instance of dentedness or rippedness with a part. And, significantly, physical wholes often share token causal powers and responsibilities with their parts, so that both can cause one effect without overdetermination. Of course not all properties of a whole are shared with its parts: you can build a circle out of square parts. But as long as unshared properties are themselves grounded in by some pattern of shared properties – as the shape of a circle is accounted for by the location properties it shares with its many square parts – then we need no 'further reality' to make the whole: it simply shares the reality of its parts.

What I call the 'experience-sharing' approach hopes that something similar is true of macrosubjects: each of their primitive experiential properties belongs both to the whole and to one of its parts, and their other experiential properties are simply patterns of these shared primitive experiences. This changes the shape of the subject-summing problem. If token experiences are not shared, we need to explain not just a new subject, but a whole new stream of consciousness. But if the whole shares token experiences with its parts, we already have our stream of consciousness. And if we take human subjects to be identical with certain physical systems (brains), we already have our candidate macrosubject. So given this composite entity, and this collection of experiences in its parts, the subject-summing problem shrinks to: why is this entity related to these experiences suitably to be (one of) their subject(s)? This does not remove the problem but might make it seem less insuperable.

Consider three challenges to experience-sharing: that it intensifies the 'easy problems' of combination, that it is incompatible with the nature of subjectivity, and that it does not explain subject-summing. The first concern is that if the whole has the very same experiences as its parts, it becomes harder to explain apparent discrepancies between human experience and what we can reasonably posit at the micro-level. For instance, the micro-level is fantastically fine-grained, but qualitatively homogeneous, while human experience is comparatively coarse-grained, but qualitatively diverse. Panpsychists who deny experience-sharing might dodge this problem by saying that macro-experience is dependent on, but still entirely distinct from, micro-experience. But no such move is available to the experience-sharer. These problems may still be soluble, but experience-sharing makes them harder.

A second concern is that experience-sharing may contradict something basic about subjecthood. For instance, we might think that conscious experience is essentially 'private', directly knowable by only one subject, in contrast to the 'public' world of physical things. Since *having* an experience seems closely linked to being able to know it directly, privacy seems to rule out experience-sharing. But perhaps the experience-sharer can say that our intuitive idea of privacy is correct, but only when applied to *discrete* (non-overlapping) subjects, rather than to *distinct* (non-identical) subjects. Since my parts are distinct from me but not discrete from me, their sharing my experiences is compatible with this qualified form of privacy (cf. Roelofs 2019: 121 ff).

A related worry, articulated in different forms by Basile (2010) and Coleman (2012, 2014), is that experience-sharing conflicts with the holistic, perspectival, character of consciousness. Even if some sort of 'element' were shared between two consciousnesses, this element will be experienced differently from their two perspectives, due to the other contents of their respective consciousnesses. And if they experience it differently, surely we should count it as a different experience: hence a single experience cannot be shared. To put it another way, for experience-sharing to explain macrosubjects, we must be able to build up a single conscious field from individual experiences with some degree of independent existence, which is what this line of objection denies. Against this, the experience-sharer could accept that consciousness is holistic, but analyse this in terms of mutual influence among elements, so that the parts are fundamental but also heavily affected by each other. This would imply that the total experience of one component subject will depend on the experiences of the *others*: its phenomenology somehow reflects theirs (see Roelofs 2016, 2019).

A third worry: even if the experience-sharing approach changes the explanandum ('why do physical composites bear the 'being-a-subject-of' relation to their parts' experiences?'), does it

actually explain that? It at least does better than physicalism. Recall section 2's distinction between explaining-what and explaining-that: experience-sharing at least accomplishes explaining-what. If we knew that the parts of me had certain experiences, and knew that I had *some* experiences, it seems we could deduce which experiences I would have – namely, those of my parts (and whatever other more complex experiences those might underlie).[4]

But why does the macrosubject have *any* experiences – why are there macrosubjects at all? If there is no contradiction in a composite of microsubjects lacking experience, we still need an explanation for why we do not inhabit a world of microsubjects without macrosubjects. One possibility is that it is somehow in the nature of *composition in general* that wholes 'inherit' properties from their parts, for at least some range of properties (this makes most sense if composition is something like identity, cf. Baxter 1988; Lewis 1991, 1993; Sider 2007; Cotnoir 2013; Baxter and Cotnoir 2014). Or it might be the nature of *subjects* that all their properties are reducible to certain patterns of resemblance, representation, or causation among experiences (as on NeoLockean accounts of personhood, cf. Parfit 1984, 1999; Shoemaker 1997). In both these cases (both explored in Roelofs 2019), the explanation would be *a priori*, but would require defending a contentious position on an independent metaphysical question.

Alternatively, the experience-sharing theorist might posit a cross-level law of nature that wholes (conditionally or unconditionally) share the experiences of their parts. This means accepting the possibility of a 'panpsychic zombie world', where experiences were not shared: we know we are not in such a world, because we know ourselves to be conscious, but this knowledge is *a posteriori*. This implies a sort of 'emergentism', in that macrosubjects do not follow *a priori* from even the fullest account of their parts. However, *a posteriori* experience-sharing lacks the theoretical vices which I discussed in section 2. It does not generate causal competition between microsubjects and macrosubjects, if causal responsibility is tied to particular shareable properties. It does not drive a wedge between mental and physical structure – indeed, it posits a law of nature to make the mental realm behave the same way as the physical realm. And because the posited law does not apply specifically to humans and other animals, it lacks the ungainly complication required to specify such a narrow range of application.

Overall, experience-sharing offers a high-risk, high-reward strategy for panpsychists. It is the approach most open to refutation – on grounds of privacy, perspectival holism, and the intensified 'easy problems'. But if these objections can be resolved, panpsychists get a better explanation than either emergentists (for simplicity, mental-physical correspondence, and causal non-competition) or physicalists (for explaining-*what* the macrosubject experiences), whether or not the fact that macrosubjects experience anything is explained by the metaphysics of subjecthood, or of composition, or by a cross-level law.

3.2. Diachronic Fusion

Could the challenges facing experience-sharing show a problem with the whole idea of two coexisting levels of experience? This thought leads some panpsychists (e.g. Seager 2010, 2017; Mørch 2014, cf. Humphreys 1997) to theorise combination as the 'fusing' of many subjects into one, so that as soon as the macrosubject exists the microsubjects are gone. There are first some parts and no whole, and then a whole and no parts, and the explanatory relation between them is diachronic, not synchronic.

Part of the appeal of the fusion approach is that in a sense there is no emergence at all, because everything goes on 'at the micro-level'. All subjects become 'micro-subjects', since none is composed of any others. The laws governing fusion, therefore, are 'micro-level' laws – or rather, they are laws operating on the most basic level, rather than 'inter-level' emergence laws. Thus human consciousness is entailed *a priori* by a full specification of 'micro-level' facts and laws, meeting our 'gold

standard' for explanation. Yet this arguably satisfies the letter while violating the spirit: the human subject may be synchronically simple but its complexity and causal powers are macro-scale, and so in our intuitive sense of 'micro-level', the fusion approach denies the adequacy of the micro-level. To properly evaluate it, we must look to the criteria outlined in section 2. Does it offer theoretical simplicity, isomorphism with physics, and causal harmony?

Seager (2010, 2017) argues that we should find fusion unproblematic, since there are examples available in physical science. His first example (also used by Humphreys 1997: 15ff) is quantum entanglement: when particles become entangled, they 'form a new state whose mathematical representation cannot be decomposed into a product of the representations of the constituents' (Seager 2017: 12). His second example is black holes, which may satisfy a 'no-hair' conjecture according to which all details of the matter which enters them are permanently abolished, leaving the hole with only three features: mass, charge, and angular momentum. Two black holes of the same mass, charge, and angular momentum are identical in all respects, lacking any 'hair' that might distinguish them.

If either of these provided a good model for the human mind, we really would have a good micro-level explanation. For in both cases, the laws governing the fusion of many entities into one are simply the general laws of physics, not laws tailored to specific complex cases. However, examining how these two examples differ reveals a dilemma for the fusion approach. Consider the problem of 'basal loss', identified by Wong in his critique of Humphreys:

> The basal properties giving rise to [an emergent property] also constitute myriad non-emergent, structural properties. . . . If these lower level properties literally ceased to be in fusing . . . then so, it seems, would those structural properties.
>
> (Wong 2006: 355)

Suppose, for instance, that certain of my neurone-states vanish into a fused mental state; what becomes of the non-emergent total neural state that they composed? If it vanishes also, it will deprive the brain of its mass, volume, shape, visible appearance, etc. Yet clearly brains do retain these non-emergent physical properties, suggesting that they retain the underlying neurone-states.

Note that for 'hairless' black holes, basal loss makes sense. What interested us in them is that the specific features of what enters them really do seem to be *lost*. This makes it both an excellent model of fusion, and a bad model for anything supposed to happen in the brain. Quantum entanglement fits the brain better, but quantum entanglement is very different from black hole formation. Entangled systems retain all the features of their 'parts'; rather than losing anything, they add something, namely lawlike relations between these features. Consequently, Wong's point about basal loss seems to count against treating entanglement as fusion.

Note that although the properties of an entangled system cannot be explained just by the individual properties of its parts, this does not mean that they cannot be explained by those individual properties, along with the 'entanglement relations' between them. These relations are certainly odd, connecting the properties of disparate objects but not behaving like other 'causal' interactions described by physics (in particular, ignoring the speed-of-light time limit). But any account of quantum phenomena must be odd, and there need be nothing offensive to reductionistic scruples about relations among parts being essential to a whole's behaviour, as long as those relations follow from the micro-level laws.

The fact that we *can* treat entangled systems as composites does not show that we cannot or should not treat them as a sort of emergent simple. So the holistic treatment might still be the *model* for the mental case. But there is something problematic about treating them as fusion, which the black hole example brings out. Consider two particles which become entangled with respect to certain of their properties (e.g. spin) but not with respect to all (e.g. charge). The resultant entangled

system still has a certain total charge and has portions of that charge at certain locations. Why does it have these charge properties? Surely because the particles that entered into it had those charge properties. But why should the properties of non-existent particles explain the properties of this system, unless they are in some sense still around? The charge was not involved in the interaction that entangled them, so it need not be mentioned in any sort of 'fusion law' of the form 'when property X and property Y interact, they form a new property Z' (see Humphreys 1997 for details of the 'fusion operation'). The entangled system might have simply *not had* any charge properties, as the example of the hairless black hole illustrates. So it seems we need extra laws, or extra clauses in the laws, saying that unless otherwise specified, the charge of an entangled system is inherited from the now-vanished particles that brought it into existence. But this seems like an unattractive multiplication of laws just to account for what seem like trivial results. It seems more efficient to see the entangled system as composed of the original particles, now differently related – or else to be thoroughly holistic, denying the fundamental existence of individual particles either before or after entanglement. But in neither of these cases is there really any fusion going on.[5]

This poses a dilemma when considering the brain. If the panpsychist's fusion law specifies that microsubjects displaying (say) a certain functional organisation fuse into a brain-sized subject, this leaves open that all their properties irrelevant to functional organisation simply vanish, leaving behind a macrosubject which 'has no hair'. Yet the brain seems to retain all the physical features of its parts, including the incredibly specific (though functionally irrelevant) distribution of features across microscopically small locations. But fundamental laws specifying that this one thing, quite distinct from the many things which go into it, should nevertheless match them exactly in trillions of respects, are, as Mørch admits, 'not very simple and elegant' (2014: 190).

We could avoid these worries about basal loss by letting mental and physical drift apart. Perhaps the experiential parts of the brain fuse into a single macrosubject, but the physical parts do not correspondingly fuse into a single physical entity (Seager 2017: 15). But this runs into causal exclusion – not between between macrosubjects and microsubjects, but between macrosubjects and micro*physics*. If the many particles account for everything the brain causes, the distinct single macrosubject faces epiphenomenalism. Certainly, it is hard to see how to retain isomorphism between mental and physical, let alone the Russellian identification of one as the categorical basis of the other.

The fusion approach deduces human subjects *a priori* from facts that can be called 'micro-level', but risks losing the structural advantages associated with panpsychism. Mental fusion without physical fusion threatens causal harmony and mental-physical isomorphism; fusion that is both mental and physical requires unparsimonious fusion laws. Perhaps the approach can somehow avoid both problems, but it is hard to see how.

3.3. Phenomenal Bonding

Micro-physical explanations typically make essential reference to relations among components – few macroscopic phenomena are explained by microphysics if we ignore causal, spatial, and other relations. So maybe the subject-summing problem results from not conceiving component subjects *as properly related*.

Of course, we can imagine subjects standing in many relations – resembling each other, acting on each other, communicating with each other, etc. But these familiar relations don't seem to help: indeed, a classic way to dramatise the subject-summing problem is to imagine human subjects so related (interacting, talking, touching, etc.) and observe how easily conceivable it is that there be no further consciousness belonging to the group (see Plotinus 1956: 346; Brentano 1987: 293; James 1890: 160). But even if familiar relations cannot do the job, perhaps some relation previously unrecognised does better. Following Goff (2017a, 2017b) I will call this relation 'phenomenal bonding'.

But what is phenomenal bonding, beyond 'the relation that solves the subject-summing problem'? Goff maintains that its nature actually precludes our understanding it, because we cannot learn about it either introspectively or perceptually. We cannot understand inter-subject relations introspectively, for introspection reveals only one subject, not many. And we cannot understand it perceptually, because it is essentially subjective, or 'inner', and perception shows us only what is objective and 'outer' (2017a: 293–294). Thus Goff admits this approach 'leads to a kind of mysterianism' (2017a: 294).[6]

So while the monadic properties of microsubjects do not explain the consciousness of macrosubjects, those properties *together with phenomenal bonding relations* do. If we had a proper conception of those relations, we would find it inconceivable that subjects standing in them *not* compose a subject. The subject-summing problem arises because we lack, and cannot acquire, a proper conception of the bonding relations.

Goff suggests we can indirectly characterise the bonding relation as the 'deep nature' of spatial relations (2017: ch. 7). This talk of 'deep natures' assumes Russellian monism: physics describes only the abstract structure of things, not their intrinsic nature. Russellian panpsychists say that the deep nature of physical properties is consciousness; Goff extends this by saying that the deep nature of physical relations is a phenomenal relation.

How to evaluate the phenomenal bonding approach? In one sense, it explains macrosubjects: it postulates something whose nature entails them, and it provides a principled reason why we cannot grasp this something. But in another sense it explains nothing: lacking a grasp of the explanatory base, we gain no illumination as to why there are macrosubjects with any particular features.

For one thing, we have no idea when and where the bonding relation obtains. Suppose we identify it with space or causation: does it hold wherever there is any degree of these (in which case it probably connects every subject in the universe to every other), or does it require some threshold of proximity or intensity, or does it itself come in degrees? This makes it hard to judge parsimony. Perhaps, given our ignorance, we should suppose the most parsimonious distribution possible, such as an entirely thorough-going one (as in Goff 2013). But this seems problematic, posing what Rosenberg (1998, 2004) calls the 'boundary problem': it is easy to see how a universally distributed bonding relation could yield a single cosmic mind, or an infinity of overlapping subjects. But why exactly do we get the specific human subjects that we wanted to explain, or at least why do they in particular have such prominence, being the only ones that get recognised as such? Perhaps this challenge can be met, but without any insight into the bonding relation, the particular size and shape of the human subject starts to seem mysterious.

Moreover, when two microsubjects get phenomenally bonded, thereby forming a composite subject, we are told basically nothing about what this subject's experience will be like. Perhaps its experiences will be type-identical to those of some or all of its parts, or perhaps token-identical, or perhaps quite different. Goff insists that *constituting* an experience need not mean *characterising* it (2017: 189 ff). But this invites the question: what *does* characterise the macrosubject's experience? The phenomenal bonding approach does not seem to tell us.

Overall, I think phenomenal bonding provides a good fall-back position for the panpsychist, if all other approaches fail. The bonding approach can claim narrow advantages both over physicalism (because while it does not *provide* an explanation of consciousness, it postulates a suitable explanatory base, as opposed to the knowably insufficient physicalist base) and over emergentism (because while it does not avoid positing an epistemically brute 'factor X', this factor is simple, general, and already operative at the micro-level). It may however have trouble showing an advantage over non-standard forms of physicalism (see Stoljar this volume), which postulate unknowable features of physical reality but not fundamental consciousness. But if the subject-summing problem cannot otherwise be solved, phenomenal bonding may be the best panpsychists can do.

4. A Combined Proposal

I think the best approach for Russellian panpsychists combines experience-sharing and phenomenal bonding. If we hold *both* that token experiences can belong to multiple subjects, *and* that the formation of macrosubjects depends crucially on distinctively phenomenal relations, we mitigate some of the shortcomings of each approach. Moreover, we get a productive direction for future research: extrapolating intra-subject phenomenal relations into inter-subject phenomenal relations.

First, experience-sharing makes phenomenal bonding less mysterian. If the bonding relations obtain between our own component subjects, and pertain specifically to their experiences, and if we ourselves share those experiences, we do have access to the relations after all. We know them as structural relations among our own experiences, such as phenomenal unification, composition of content, mutual reference or accessibility, or whatever else phenomenology reveals. On this approach, Goff is wrong to say that introspection reveals only one subject: introspection shows us at least some of our component subjects as well as ourselves. Second, phenomenal bonding helps stop experience-sharing from 'making the easy problems harder'. If experiences change each other's phenomenal character through being bonded, and if moreover not all parts of me, or not all of their experiences, are bonded, that might explain why our experience is sensitive to certain features of our brain and not others.

Neither of these advantages is a completed result: the easy problems are still quite hard, and we still need a fuller account of the bonding relations. But this combined proposal suggests a methodology to address these challenges. First, find distinctively phenomenal relations within human experience. Then, evaluate them as potential bonding relations, asking if they are suitably basic and explanatory. If they seem promising, work out how those relations could hold inter-subjectively, between experiences of distinct subjects. This means both identifying any problems that follow from such a supposition and finding points of contact between our chosen intra-subjective relations and our inter-subjective relations with each other. This methodology, which I pursue in other work (Roelofs 2016, 2017, 2019), is not guaranteed to succeed: but hopefully it will be illuminating, whether it succeeds or fails.

Notes

1. When I speak of 'emergentists', I mean 'strong' emergentism, on which some property of a complex could not be predicted from its emergence base even with the most complete knowledge of the relevant properties of that base. Instead, it arises from a *sui generis* law. Moreover, I will assume that (strong) emergentists are not panpsychists. 'Weak' emergentists, who hold emergent phenomena to be predictable in principle, even if not in practice, I count as physicalists or panpsychists, depending on the features of their emergence base (see Broad 1925; Chalmers 2006; Wilson 2016).
2. See James (1890: 147–61), Nagel (1986: 50), Van Cleve (1990: 219), Rosenberg (1998, 2004), Strawson (2006), Goff (2006, 2009a, 2009b), Basile (2010), Shani (2010), Seager (2010, 2017), Coleman (2012, 2014), Sebastián (2015), Mørch (2014), Roelofs 2014, 2016, 2019), and Mendelovici (this volume).
3. Of course, not all panpsychists rest their position on the explanatory gap faced by physicalism. There are other arguments for panpsychism available, most especially the 'intrinsic natures' argument that seems to show all non-panpsychist views to be unparsimonious and borderline unintelligible (e.g. Seager 2006; Coleman 2009). But since much of the recent interest in panpsychism does derive from explanatory concerns, we should consider what panpsychists have to do to maintain their claim of explanatory superiority.
4. Note that in order to deduce this we would have to rule out that I had some other experiences, unrelated to those of my parts. But we are perfectly entitled to rule this out, according to the most prominent models of how explanatory deductions proceed in the physical realm (Chalmers and Jackson 2001; Chalmers 2012), which explicitly requires a 'that's-all' clause in any micro-level explanation.
5. Will it help to say that the entangled system is not really *partless*, but simply holistic in the sense of being more fundamental than its parts (cf. Mørch 2014: 167–75, 191)? We can still say that at the fundamental level, the parts are replaced by a simple whole. And it is not clear that this provides any *genuine* persistence of the parts,

because it is not clear whether there can be genuine identity between a fundamental entity at one time and a non-fundamental entity at a later time.

6. Chalmers (2017), following Dainton (2011), offers a less mysterian proposal: that the phenomenal bonding relation is 'co-consciousness', the relation studied in the literature of the unity of consciousness. I think this version of the bonding approach is stronger than the mysterian one, and have defended a version of it myself (Roelofs 2016, 2019).

References

Albahari, M. (this volume). 'Abhidharma Panprotopsychist Metaphysics of Consciousness'. In W. Seager (ed.), *The Routledge Handbook of Panpsychism*. London: Routledge.

Alter, T., and Nagasawa, Y. (eds.) (2015). *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford: Oxford University Press.

Basile, P. (2010). 'It Must Be True – But How Can It Be? Some Remarks on Panpsychism and Mental Composition'. *Royal Institute of Philosophy Supplement*, 67: 93–112.

Baxter, D. (1988). 'Many-One Identity'. *Philosophical Papers*, 17 (3): 193–216.

Baxter, D., and Cotnoir, A. (eds.) (2014). *Composition as Identity*. Oxford: Oxford University Press.

Blackmore, S. (2002). 'There Is No Stream of Consciousness'. *Journal of Consciousness Studies*, 9 (5–6): 17–28.

Blamauer, M. (this volume). 'The Crux of Subjectivity the Subjective Dimension of Consciousness and Its Role in the Arguments for and Against Panpsychism'. In W. Seager (ed.), *The Routledge Handbook of Panpsychism*. London: Routledge.

Block, N., and Stalnaker, R. (1999). 'Conceptual Analysis, Dualism, and the Explanatory Gap'. *Philosophical Review*, 108: 1–46.

Brentano, F. (1987). *The Existence of God: Lectures Given at the Universities of Worzburg and Vienna, 1868–1891*. Ed. and Trans. S. Krantz. Nijhoff International Philosophy Series. Dordrecht: Nijhoff.

Broad, C. D. (1925). *The Mind and Its Place in Nature*. London: Kegan Paul.

Chadha, M. (this volume). 'Abhidharma Panprotopsychist Metaphysics of Consciousness'. In W. Seager (ed.), *The Routledge Handbook of Panpsychism*. London: Routledge.

Chalmers, D. (1995). 'Facing Up to the Problem of Consciousness'. *Journal of Consciousness Studies*, 2 (3): 200–19.

Chalmers, D. (2006). 'Strong and Weak Emergence'. In P. Clayton and P. Davies (eds.), *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*. Oxford: Oxford University Press, pp. 244–56.

Chalmers, D. (2012). *Constructing the World*. Oxford: Oxford University Press.

Chalmers, D. (2017). 'The Combination Problem for Panpsychism'. In G. Brüntrup and L. Jaskolla (eds.), *Panpsychism: Contemporary Perspectives*. Oxford: Oxford University Press, pp. 179–214.

Chalmers, D., and Jackson, F. (2001). 'Conceptual Analysis and Reductive Explanation'. *The Philosophical Review*, 110 (13): 315–61.

Coleman, S. (2009). 'Mind Under Matter'. In D. Skrbina (ed.), *Mind That Abides*. Amsterdam: Benjamins, pp. 83–107.

Coleman, S. (2012). 'Mental Chemistry: Combination for Panpsychists'. *Dialectica*, 66 (1): 137–66.

Coleman, S. (2014). 'The Real Combination Problem: Panpsychism, Micro-Subjects, and Emergence'. *Erkenntnis*, 79: 19–44.

Coleman, S. (2017). 'Panpsychism and Neutral Monism: How to Make Up One's Mind'. In G. Brüntrup and L. Jaskolla (eds.), *Panpsychism: Contemporary Perspectives*. Oxford: Oxford University Press, pp. 249–82.

Cotnoir, A. (2013). 'Composition as General Identity'. In K. Bennett and D. Zimmerman (eds.), *Oxford Studies in Metaphysics*, vol. 8. Oxford: Oxford University Press, pp. 295–322.

Dainton, B. (2011). 'Review of *Consciousness and Its Place in Nature*'. *Philosophy and Phenomenological Research*, 83 (1): 238–61.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown and Co.

Diaz-Leon, E. (2011). 'Reductive Explanation, Concepts, and a Priori Entailment'. *Philosophical Studies*, 155: 99–116.

Eddington, A. (1929). *The Nature of the Physical World*. Cambridge: Cambridge University Press [1927 Gifford Lectures. First edition published in 1928].

Feigl, H. (1958). 'The "Mental" and the "Physical"'. In H. Feigl, M. Scriven, and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science, vol. 2: Concepts, Theories and the Mind-Body Problem*, Minneapolis: University of Minnesota Press, pp. 370–497.

Goff, P. (2006). 'Experiences Don't Sum'. *Journal of Consciousness Studies*, 13 (10–11): 53–61.

Goff, P. (2009a). 'Why Panpsychism Doesn't Help Us Explain Consciousness'. *Dialectica*, 63 (3): 289–311.

Goff, P. (2009b). 'Can the Panpsychist Get Round the Combination Problem?' In D. Skrbina (ed.), *Mind That Abides: Panpsychism in the New Millennium*. Amsterdam: John Benjamins.
Goff, P. (2012). 'There Is More Than One Thing'. In P. Goff (ed.), *Spinoza on Monism*. New York: Macmillan, pp. 113–22.
Goff, P. (2013). 'Orthodox Property Dualism + Linguistic Theory of Vagueness = Panpsychism'. In R. Brown (ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*. Dordrecht: Springer, pp. 75–91.
Goff, P. (2017a). 'The Phenomenal Bonding Solution to the Combination Problem'. In G. Brüntrup and L. Jaskolla (eds.), *Panpsychism: Contemporary Perspectives*. Oxford: Oxford University Press, pp. 283–304.
Goff, P. (2017b). *Consciousness and Fundamental Reality*. Oxford: Oxford University Press.
Goff, P. (this volume). 'From Russellian Monism to Cosmopsychism'. In W. Seager (ed.), *The Routledge Handbook of Panpsychism*. London: Routledge.
Humphreys, N. (1997). 'How Properties Emerge'. *Philosophy of Science*, 64: 1–17.
Huxley, T. H. (1986). *Lessons in Elementary Physiology*. London: Macmillan (originally published 1866).
James, W. (1890). *The Principles of Psychology*. Cambridge, MA: Harvard University Press.
Jaskolla, L., and Buck, A. (2012). 'Does Panexperiential Holism Solve the Combination Problem?' *Journal of Consciousness Studies*, 19: 190–9.
Levine, J. (1983). 'Materialism and Qualia: The Explanatory Gap'. *Pacific Philosophical Quarterly*, 64: 354–61.
Lewis, D. (1991). *Parts of Classes*. Oxford: Blackwell.
Lewis, D. (1993). 'Many, But Almost One'. In J. Bacon, K. Campbell and L. Reinhardt (eds.), *Ontology, Causality, and Mind: Essays in Honour of D. M. Armstrong*. Cambridge: Cambridge University Press, pp. 23–45.
Loar, B. (1990). 'Phenomenal States'. *Philosophical Perspectives*, 4: 81–108.
Lockwood, M. (1981). 'What Was Russell's Neutral Monism?' *Midwest Studies in Philosophy (The Foundations of Analytic Philosophy)*, VI: 143–58.
McQueen, K. (2015). 'Mass Additivity and A Priori Entailment'. *Synthese*, 192 (5): 1373–1392.
Mendelovici, A. (this volume). 'Panpsychism's Combination Problem Is a Problem for Everyone'. In W. Seager (ed.), *The Routledge Handbook of Panpsychism*. London: Routledge.
Mizrahi, V. (2009). 'Is Colour Composition Phenomenal?' In D. Skusevich and P. Matikas (eds.), *Color Perception: Physiology, Processes and Analysis*. Hauppauge: Nova Science Publishers, pp. 185–202.
Mørch, H. H. (2014). *Panpsychism and Causation: A New Argument and a Solution to the Combination Problem*. Doctoral Dissertation, University of Oslo. https://philpapers.org/archive/HASPAC-2.pdf.
Nagel, T. (1979). 'Panpsychism'. In T. Nagel (ed.), *Mortal Questions*. Cambridge: Cambridge University Press, pp 181–195.
Nagel, T. (1986). *The View from Nowhere*. Oxford: Oxford University Press.
Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
Parfit, D. (1999). 'Experiences, Subjects, and, Conceptual Schemes'. *Philosophical Topics*, 26: 217–70.
Plotinus. (1956). *Enneads*. Ed. and Trans. S. MacKenna and B. Page. London: Faber and Faber. Accessed July 2013. http://archive.org/stream/plotinustheennea033190mbp#page/n11/mode/2up.
Roelofs, L. (2014). 'Phenomenal Blending and the Palette Problem'. *Thought*, 3 (1): 59–70.
Roelofs, L. (2016). 'The Unity of Consciousness, Within and Between Subjects'. *Philosophical Studies*, 173 (12): 3199–221.
Roelofs, L. (2017). "Rational Agency without Self-Knowledge: Could 'We' Replace 'I'?" *Dialectica* 71 (1): 3–33.
Roelofs, L. (2019). *Combining Minds*. New York: Oxford University Press.
Rosenberg, G. (1998). 'The Boundary Problem for Phenomenal Individuals'. In S. Hameroff, A. Kaszniak, and A. Scott (eds.), *Toward a Science of Consciousness: The First Tucson Discussions and Debates (Complex Adaptive Systems)*. Cambridge, MA: MIT Press, pp. 149–56.
Rosenberg, G. (2004). *A Place for Consciousness: Probing the Deep Structure of the Natural World*. Oxford: Oxford University Press.
Russell, B. (1927). *The Analysis of Matter*. London: Kegan Paul.
Seager, W. (1995). 'Consciousness, Information and Panpsychism'. *Journal of Consciousness Studies*, 2–3: 272–88.
Seager, W. (2006). 'The Intrinsic Nature Argument for Panpsychism'. *Journal of Consciousness Studies*, 13 (10–11): 129–45.
Seager, W. (2010). 'Panpsychism, Aggregation and Combinatorial Infusion'. *Mind and Matter*, 8 (2): 167–84.
Seager, W. (2017). 'Panpsychist Infusion'. In G. Brüntrup and L. Jaskolla (eds.), *Panpsychism: Contemporary Perspectives*. Oxford: Oxford University Press, pp. 229–48.
Sebastián, M. A. (2015). "What panpsychists should reject: on the incompatibility of panpsychism and organizational invariantism." *Philosophical Studies* 172 (7):1833–1846.

Shani, I. (2010). "Mind Stuffed with Red Herrings: Why William James' Critique of the Mind-Stuff Theory Does Not Substantiate a Combination Problem for Panpsychism." *Acta Analytica* 25 (4): 413–434.
Shani, I. (2015). 'Cosmopsychism: A Holistic Approach to the Metaphysics of Experience'. *Philosophical Papers*, 44 (3): 389–437.
Shoemaker, S. (1975). 'Functionalism and Qualia'. *Philosophical Studies*, 27: 291–315.
Shoemaker, S. (1997). 'Parfit on Identity'. In J. Dancy (ed.), *Reading Parfit*. Oxford: Blackwell, pp. 135–48.
Sider, T. (2007). 'Parthood'. *Philosophical Review*, 116: 51–91.
Smart, J. J. C. (1959). 'Sensations and Brain Processes'. *The Philosophical Review*, 68 (2): 141–56.
Stoljar, D. (2014). 'Four Kinds of Russellian Monism'. In U. Kriegel (ed.), *Current Controversies in Philosophy of Mind*. London: Routledge, pp. 17–39.
Stoljar, D. (this volume). 'Panpsychism and Nonstandard Materialism: Some Comparative Remarks'. In W. Seager (ed.), *The Routledge Handbook of Panpsychism*. London: Routledge.
Strawson, G. (2006). 'Realistic Monism: Why Physicalism Entails Panpsychism'. In A. Freeman (ed.), *Consciousness and Its Place in Nature*. Exeter: Imprint Academic, pp. 3–31.
Van Cleve, J. (1990). "Mind-Dust or Magic? Panpsychism versus Emergence." *Philosophical Perspectives* 4: 215–226.
Wilson, J. (2016). 'Metaphysical Emergence: Weak and Strong'. In T. Bigaj and C. Wüthrich (eds.), *Metaphysics in Contemporary Physics* (Poznan Studies in the Philosophy of the Sciences and the Humanities). Leiden: Brill Rodopi, pp. 251–306.
Wong, H. Y. (2006). "Emergents from Fusion." *Philosophy of Science*, 73: 345–367.