

Combination Problems and Combination Solutions

The combination problem is not the most widespread objection to panpsychism, but it is the most serious. Arguably the most widespread is simply an incredulous stare, but the combination problem is more serious because it challenges panpsychism on its own terms: most panpsychists reject physicalism because it seems unable to explain human consciousness, so it would be a serious problem if panpsychism couldn't either. And that is what the combination problem alleges: that for various reasons, conscious minds don't combine - don't form wholes which are also conscious, and whose consciousness is fully explained by that of their parts, and by how those parts relate.

I've done a lot of writing about the combination problem - or, as I prefer, combination problems, since there are a remarkable diversity of issues that get brought up here, such that when someone mentions 'the combination problem', you often have little idea what issues they really have in mind. I think the problems are solvable: to put it more positively, I think minds can combine. But none of them are simple - often, different solutions make sense depending on where one stands on certain other disputed issues, so that no single account will be satisfying to everyone, and other times the solutions to one problem only make sense in relation to solutions to other problems.

Indeed, a conception of minds as non-combining things - as simple, unified, isolated, things that can never be analysed into components minds - is deeply rooted in our normal ways of thinking. Of course, most of what we have learnt from neuroscience, psychology, and cognitive science, is in tension with this conception (what I call the Anti-Combination intuition),

and many writers have challenged it and tried to chip away at it. Panpsychism simply radicalises this existing trend.

In this paper I want to summarise the account of panpsychist mind-combining that I have suggested and defended in a few papers and in my recent book *Combining Minds*. It is deliberately quick and sketchy, and the arguments are more gestured at than made, but I hope it will serve to convey in outline why I do not think the combination problems provide a successful objection to panpsychism.

Section 1: What Combines, When and Why?

Let me first make a few remarks about the overall ‘shape’ of the combinationism I’ll defend. The usual focus in this debate is on the question ‘how does human consciousness arise from the combining of fundamental consciousness?’, but this risks obscuring broader questions about when (and where, and how) consciousness combines. To the first approximation, the candidate answers are ‘always’, ‘never’, and ‘sometimes’. I am attracted towards the first option: ‘universalism’ about combination (aka ‘unrestricted phenomenal composition’). I prefer it over the ‘never’ option because I know that consciousness has combined at least once (in my head), and I prefer it over the ‘sometimes’ option because I can’t see how any intuitive and natural restriction on combination could be made precise. That is, between any system which intuitively ought to be conscious (e.g. me) and any system that intuitively ought not to (e.g. a pile of dirt), there are only differences of degree, out of which a spectrum of intermediate cases could be

constructed so that any precise line drawn would seem arbitrary (for a similar argument see Chalmers 1996, p.297, Goff 2013, Mørch 2014, pp.153-154).

This very quick argument obviously relies on the premise that consciousness is not vague: it is either present or absent, not indeterminate between the two. I think that must be correct if consciousness is a truly fundamental property, something without analysis into more basic properties. Vagueness is just imprecision in our concepts, and doesn't make sense for the world's fundamental constituents. But vagueness *does* make sense for many of the things we might mean by 'consciousness': cognitive access, sensory integration, coordinated action, alertness, attention, reflective thought. That is, insofar as these are complex functional capacities, it makes sense to think there will be only a fuzzy boundary between systems that have them and systems that don't. And universalism seems clearly wrong for most of these capacities: even if a pile of dirt is conscious, it's not organised the right way to be an agent, a cognizer, a 'mind'.

So the basic shape of my combinationism relies on distinguishing 'consciousness' as bare phenomenal consciousness, the fundamental raw feels that are inherent in matter itself, and 'consciousness' as cognition and agency, the organisation that distinguishes my mind from a pile of dirt. The former combines always and everywhere; the second combines only when the right causal structure arises through evolution or artificial design. I'll refer to beings with both sorts of 'consciousness' as 'intelligent subjects', and beings with only the first as 'mindless subjects'.¹

Another key question is whether different constituted minds overlap with one another. Some panpsychists (and others) want to avoid such overlap, and posit various sorts of

¹ A being with the second sort but not the first would be a P-zombie; I think it very unlikely that any P-zombie-type systems exist in our universe.

‘exclusion’ or ‘anti-nesting’ principles such that the consciousness of, say, a human being somehow ensures that no part of it, nor any whole it enters into, nor any system that shares parts with it, is conscious at the same time (see for example Putnam 2003, 215–216, Tononi 2012, p.59–68, Kammerer 2015, Mørch 2019, p.135, cf. Schwitzgebel 2015, Goff and Roelofs Forthcoming).

Such posits seem to me unnecessary, inelegant, absurd in some of their implications, and markedly out of step with the way nature seems to work in every other domain. So instead I embrace overlap: sitting here in this chair is not just me, and the particle-minds that ultimately compose me, but all sorts of intermediate conscious entities like my neurones, my hemispheres, my brain, my lower half, me-and-the-air-in-this-room, etc. Some of these are intelligent subjects (e.g. my brain), some are unintelligent subjects (e.g. my lower half, if the enteric nervous system doesn’t qualify as intelligent), some are something in between (e.g. my left half has some intelligent elements, like my left cerebral hemisphere, but they are cognitively connected to things outside it rather than to other parts of it).

Crucially, these entities do not ‘add up’ into some sort of unwieldy profusion: they share the same experiences (more precisely, my brain and I share all our experiences, each of my neurones shares some element of my total experience, etc.) because they are really just the same conscious matter, the same reality, carved up differently. Recognising them as all co-existing is no different from recognising that I, my left half, my cells, etc. all share the same space, the same mass, the same causal powers, because we are fundamentally the same portion of reality, carved up differently.

I think it can be useful to look at the combination problems not just as ‘how do we get human consciousness out of particle consciousness?’ but also as a broader challenge to think about consciousness as something that works like this - as an expanse that can be carved up in various ways while remaining the same, rather than as a property only of discrete, separate, isolated beings.² But this is a good sort of challenge to face, since it means thinking about consciousness the same way we think about everything else fundamental in nature.

Section 2: The Unity Problem

So let’s get onto the problems themselves. Chalmers, in his paper ‘The Combination Problem for Panpsychism’ (2017), makes a helpful division into problems of subject-combination, quality-combination, and structure-combination. The next three sections will consider those three sub-problems, but first I want to address a problem Chalmers mentions only in passing, the ‘unity problem’: how do microexperiences come together to yield a unified consciousness?

It’s common for discussions of ‘the unity of consciousness’ to begin with some gestures towards a familiar phenomenon - that not only does each of us experience many things at once, we in some sense ‘experience them together’, within a single ‘phenomenal field’ - and then to start enumerating a variety of relations that experiences can stand in and that seem to have something to do with this familiar impression of unity (see, e.g., Bayne and Chalmers 2003, Tye 2003, Bayne 2010, Schechter 2013). This variety of relations prominently includes causal ones

² This also, for me, reduces the pressingness of questions about the fundamental level - about cosmopsychism vs. micropsychism, or panpsychism vs. panprotopsychism (see e.g. Shani 2015, Coleman 2017). What interests me is thinking about how consciousness combines between any two levels, or more broadly between any two schemes of division, and that can be largely independent of what the most fundamental level of reality looks like.

(e.g. one experience may draw attention away from, or direct attention to, another, or two contrasting experiences may lead to the thought ‘those two are different!’) and representational ones (e.g. two visual experiences may present their objects as being at a certain distance from one another, or an emotion may be experienced as directed at the same object as a perception); it also includes ‘phenomenal unity’, the basic ‘co-consciousness’ or ‘experienced-together-ness’ which, unlike the above, seems to be entirely pervasive in normal human experience, and which seems like a precondition for more specific causal and representational sorts of unity.

If the unity problem is the problem of explaining these different relations, which ones? And why is it a problem? For instance, is there a special problem with explaining the causal relations that ‘unify my consciousness’? I don’t think there should be. The obtaining of certain causal relations is in general a matter for scientific explanation: the informative answer to why certain of my experiences causally influence other of my experiences will be given by some combination of neuroscience, developmental psychology, and evolutionary biology. There is of course an interesting question of how panpsychists should think about physical causation vs. mental causation (see e.g. Howell 2015, Coleman and Alter 2020), but if they have a good story about that (which I think they do), there won’t then be a special problem of explaining the exceedingly complex causal structure of the human mind/brain. And while there are intricate puzzles about explaining how experiences can have representational content, there isn’t a special puzzle about why the experiences associated with one brain are so often representationally connected - for whatever it is that explains complex content, it clearly has something to do with the causal-informational structure of the brain.

What generates a difficult ‘unity problem’, I think, is phenomenal unity. To see why there’s a problem, observe that with other sorts of unity, like complex causal interactions, we have no trouble making sense of two different subjects’ experiences standing in that relation. My experiences can interact with yours in all sorts of ways: indeed, in the right context it’s common for multiple people’s experiences to interact in such a way that they collectively approximate some of the functional structure of an intelligent agent (we call this ‘collective agency’). But when it comes to phenomenal unity, it’s far harder to make sense of distinct subjects’ experiences being related that way. Indeed, I think people sometimes have trouble even keeping clear what it means to speak of, e.g. my experiences and yours being phenomenally unified. We can easily lapse into thinking it means that because they’re all unified, they’ve become the same thing, and so we would share all experiences. It doesn’t mean that: it means I have my experiences, you have yours, and they stand in the relation we’ve called ‘phenomenal unity’, which we each first identified among our own experiences. But making sense of that is still difficult, and I think that’s what the ‘unity problem’ really comes down to: a difficulty with seeing phenomenal unity as a between-subjects relation.

So why is it hard to see phenomenal unity as an inter-subject relation? I think there are a few reasons, which I go over in my 2016 paper ‘The Unity of Consciousness, Within and Between Subjects’, but here I’ll focus on one that I didn’t discuss there but have come to think is important. It relates to the clash between what I call ‘reductionist’ and ‘primitivist’ intuitions about phenomenal unity. Reductionism here is the idea that phenomenal unity is really nothing over and above some sufficient amount of the other sorts of unity, i.e. various sorts of functional and causal connection (see, e.g., Shoemaker 2003). When experiences are interacting closely

enough, richly enough, sensitively enough, the thought goes, phenomenal unity comes for free: there's no further special ingredient needed. Primitivism is the opposite idea: that phenomenal unity is a distinct relation, not analytically connected with other sorts of unity (see, e.g. Dainton 2000). I think both reductionism and primitivism are somewhat widespread and somewhat intuitive, but they put opposite and incompatible requirements on an account of between-subjects phenomenal unity.

Reductionism suggests that it should come from the same sort of thing that gives us causal and representational sorts of unity between subjects: information-exchanging causal interactions among different conscious minds. Maybe we need a lot more of that - faster interaction, deeper interaction, richer interaction - before two human minds could be unified. But fundamentally it's just more of the same. Phenomenal unity, like intelligent subjects, will be rare but non-fundamental. Indeed, it's natural to think, as a reductionist, that there might be only a vague boundary between which forms of interactions are enough to yield phenomenal unity and which are not.

But to a primitivism, this is leaving out the central issue. If primitivism is true, then ever-more complex causal structures just aren't the right kind of thing to give you phenomenal unity. After all, consider the sorts of causal links we already have among human minds - my experiences 'interact with yours' through, e.g., us having a conversation. This interaction is strikingly indirect - my experience only affects yours through a chain of causal intermediaries. Of course, interactions within my brain also rely on causal intermediaries, but I think there's an intuition that any kind of interaction your experiences can have with mine is sharply different from interactions among my experiences, precisely because in the latter case the experiences are

already phenomenally unified. Phenomenal unity has to already be there for causal interactions to be of the right sort: it won't come about just through adding more of them. Indeed, it seems like two subjects could interact in whatever complex pattern you like, while their experiences remain nevertheless completely phenomenally separate. Isn't that at least conceivable?

To satisfy primitivist intuitions, I think a panpsychist account of phenomenal unity has to look very different from how it has to look to satisfy reductionist intuitions. Rather than saying 'just add more of the right sort of causal interaction', it should posit that some fundamental physical relation is, as a brute natural fact, phenomenal unity. This allows that it's conceivable, even possible, for two experiences to interact in any way you like, without being phenomenally unified: that is would happen in a world where Descartes was right, and each human mind was an utterly discrete immaterial thing, with no relations among minds except as mediated by strictly physical causation.

Which physical relation in our world might be phenomenal unity? It might be some particular force, or quantum entanglement, or something about sharing a space, but in practice it doesn't matter much which fundamental physical relation it is, because they are all so widespread as to be effectively all-pervasive. In my book, I express this as the 'Micro-Unity Hypothesis':

Micro-Unity Hypothesis (MUH): The inner nature of one, some, or all of the fundamental physical relations is phenomenal unity; when two microsubjects are related in the relevant way, their experiences become unified, establishing a composite experience that subsumes them. (2019, p.80)

This implies that effectively all experiences are phenomenally unified already, including mine with yours, and the differences are just regarding the richer sorts of unity. I think primitivism is true, and the Micro-Unity Hypothesis is true, but I can still recognise that this position will seem deeply weird to the extent that we have reductionist intuitions telling us that phenomenal unity automatically goes with rich causal interconnections. My experiences don't have rich causal interconnections with every other experience in the world, so how can it make sense to say they're all phenomenally unified? But insofar as we're feeling the pull of this sort of intuition, we should prefer the reductionist account given earlier: experiences get unified when they interact in complex enough ways - and that account won't be satisfying insofar as we feel the pull of primitivist intuitions.

The upshot is that no single panpsychist account of 'the unity of consciousness' will ever seem fully satisfying. We need two accounts, working in tandem, each capturing some of our intuitions. In *Combining Minds* I try to develop both of those two accounts, and ultimately I'm happy for people to accept either one. Though I lean towards the primitivist side, I'm more concerned to bring out and illuminate the fact that a choice has to be made, and that it reflects an independent, non-panpsychism-related choice about how to think of phenomenal unity itself.

Section 3: The Subject-Combination Problem

A second sub-problem, and the one that has perhaps received the greatest attention, is the subject-combination or subject-summing problem. Granting experiences to fundamental particles can explain why there's consciousness happening here in my brain (it's made of particles, after

all), but what explains any of it being *my* consciousness, given that I am not a fundamental particle? How do we get human subjects out of microscopic subjects?

Subsection 3.1: Reductionism and Primitivism about Subjects

What I say here will somewhat echo what I said about the unity problem: how we solve the subject-combination problem depends entirely on how we conceive of subjects, and that is neither simple nor uncontentious. One philosophically influential view, for example, is the NeoLockean analysis, sometimes called the ‘psychological continuity theory’. On this view, subjects arise simply by experiences, or other mental states, standing in the right pattern of psychological relations (see, e.g. Locke 1836, Lewis 1976, Parfit 1984, Rovane 1998). What it means for me to be a subject, for example, is that all my current experiences are linked to each other, and to certain past experiences, and (unless I die right now) to certain future experiences. The links are partly causal (the way I think today is heavily affected by how I thought yesterday, and will influence how I think tomorrow), partly about resemblance (my beliefs today are largely the same as my beliefs yesterday, and will only change a little by tomorrow), and partly representational (my current memories represent my past experiences, and my current intentions represent my future volitions). While different writers emphasise different specific relations, the basic idea is that subjects are not fundamental: they are a certain sort of ‘bundling together’ of experiences.

A slightly different (though perhaps in the end compatible) analysis is that provided by various sorts of ‘functionalism’ about the mind: roughly, a subject is a system which has a

certain sort of functional structure, a certain sort of ability to process information and move between internal states in response to external stimuli (see, e.g. Shoemaker 1997, Schechter 2015). This conception differs from the NeoLockean one in being more focused on what is common to all subjects (e.g. the capacity to perceive) than on what distinguishes one subject from another (e.g. my distinctive memories), and in formulating its requirements in terms of how states or parts of a system interact, rather than how experiences or mental states relate. But both conceptions agree that the existence of a subject requires some rich sort of structure, and that this structure is to be clarified by analysing our idea of what it means to be a subject. We may call analyses like this ‘reductionist’ about subjects. In the terms I suggested earlier, these are essentially concerned with ‘intelligent subjects’, and quite possibly imply the impossibility of ‘mindless subjects’.

What is noteworthy about both of these analyses, and other reductionist ones, is what they imply about the subject-summing question, ‘how do many subjects come together to form a composite subject?’ They make this a very significant question in particular cases, because they imply that the formation of a composite subject is in no way guaranteed: it is very easy for a bunch of subjects to form a whole which is not a subject, because it lacks the right structure. Any set of randomly selected human beings will provide an example. Yet at the same time, reductionist analyses make the subject-summing question *easy* to answer in the abstract. How do many subjects come together to form a composite subject? By getting into the structure which, according to the reductive analysis, is essential to subjecthood. And while doing so may be difficult and fiddly, it is not theoretically problematic - indeed, it is generally a selling point for analyses like this that 1) human brains and bodies evidently *do* possess the right kind of

structure, and 2) the analysis is quite agnostic about the metaphysical underpinnings of that structure - it is compatible with physicalism, dualism, idealism, etc.

What this means is that if panpsychists adopt any of the popular reductionist analyses of subjecthood, there is no pressing subject-combination problem for them to address. Their theory of the micro-level says that there are a bunch of experiences here in my body; various scientific theories (molecular, physiological, developmental, etc.) explain why various assemblages of those experiences interact with one another in complex and intelligent ways. According to the reductionist analyses of subjecthood, that is *all there is* to a composite subject (me) existing. There are, admittedly, many interesting questions raised by such analyses - can fundamental particles properly be called 'subjects'? Does the formation of a composite subject necessarily submerge the subjecthood of its parts? How, in general, do component subjects relate to composite subjects? All of these are questions I explore in depth in *Combining Minds*, but none of them threaten panpsychism itself with an explanatory gap.

But let's not celebrate prematurely. Obviously a lot of panpsychists and their critics *do* see a pressing subject-combination problem. They seem to be able to conceive of all those micro-experiences occurring in my body, interacting in all the specified ways, and yet there being no such subject as 'me' experiencing any of them (see, e.g. Goff 2009, cf. Strawson 2009). I think this is a clear sign that a lot of panpsychists do not endorse any reductionist analysis of subjecthood: they think of 'subject of experience' as something like a primitive, not analysable in any more basic terms. For there to be a subject is not for some complex structure to be instantiated by some more basic things, but simply... for there to be something that experiences.

It might lack memory, perception, reflective thought, and so on: it needs only to be a ‘locus of consciousness’, a thing such that consciousness happens there.

I think this ‘primitivist’ view of subjects captures a lot of our intuitions about subjects, but I think the reductionist analyses also capture a lot of our intuitions. It is a noteworthy fact that multiple directly conflicting conceptions can capture many of our deepest intuitions about subjecthood. I think this is a major reason why debates about personal identity are as intractable and perplexing as they are. But what it means for panpsychists is that they can’t, and shouldn’t try to, answer the subject-summing question *simpliciter*, in a way that will do justice to all our intuitions about what it is to be a subject. Nobody can do justice to all those intuitions within a single account. The best that can be done is to give two accounts: one that accepts some sort of reductionist analysis - on which, I claim, there is no problem of subject-combination - and one that embraces primitivism.

So let us embrace primitivism, and reject the intuitions captured by the reductionist analyses. Let us reject, in particular, any intuition that speaking of a subject of experience necessarily means speaking of some complex cognitive capacities, or some special psychological structure that can reliably distinguish each subject from all others. I think when we scrupulously reject these intuitions, we remove all impediments to accepting an account of subject-combination that cuts the Gordian Knot, and simply says: subjects always, automatically, combine. For any set of subjects, they are collectively also a subject - to put it another way, any set of loci of consciousness is itself a locus of the very same consciousness. In *Combining Minds*, I formulated this idea as two theses, ‘Experience Inheritance’ and ‘Experience Sharing’. The first

says that whenever a part of something instantiates an experiential property, that thing does too - it 'inherits' that property from its part:

Experience Inheritance (EI): Whenever a part of aggregate x undergoes an experience (instantiates an experiential property), x undergoes that same experience. (2019, p.79)

The second adds that it's the very same token experience: the whole doesn't duplicate the part's experience, but shares in it. This requires qualifying the seeming truism that experiences are 'private', belonging only to a single subject: in its place I propose:

Weak Privacy: An experience of one subject cannot belong to any other subject not overlapping with the first. (2019, p.63)

This allows for experiences to be shared between a whole and its parts, but still holds onto the idea that two completely separate beings must have their own separate experiences.

This amounts to treating consciousness as 'aggregative' (cf. Seager 2010), just as we are used to treating fundamental properties. A collection of things with mass also has mass - not extra mass, but simply the same mass as those things have. A collection of things with volume also has volume - it occupies exactly the space which those things occupy. Of course this would be absurd if we thought of subjecthood in a way captured by the reductionist analyses - as something that essentially requires integrated cognition. For this implies that the whole formed by me and the chair I am sitting on is a subject, undergoing all of my experiences together with those of the chair. Indeed, the whole formed by any group of randomly selected people is also a conscious subject, undergoing all of their experiences at once. Yet clearly such wholes do not

exhibit integrated cognition: they are not ‘intelligent subjects.’ But if we wish to capture those intuitions, we would switch to our other account: we would say ‘well, that group of randomly selected people is not a subject *in the reductionist’s sense*, because it doesn’t have the right structure’. All sorts of random aggregates are subjects, but mostly ‘mindless subjects’, not intelligent subjects, and our everyday intuitive thinking about subjects is inconsistent about which of those it means.

Subsection 3.2: Explaining Inheritance

There is still the question of *why* subjects in the primitivist sense should freely combine in this aggregative way. There are roughly two candidate answers: natural law or composition-as-identity. The first would say that it’s just one of the basic brute facts about our universe that subjects combine like this (cf. ‘phenomenal bonding’, Goff 2017-b, Miller 2018). Perhaps it has something to do with some other basic facts about the universe, like the nature of spacetime, or the Micro-Unity Hypothesis I argued for above. In particular, a fairly simple argument can be made from the Micro-Unity Hypothesis to something close to Experience Inheritance:

Premise 1: The deep nature of one of the fundamental physical relations is phenomenal unity (the Micro-Unity Hypothesis).

Premise 2: All of the fundamental physical relations hold either universally or nearly universally.

Premise 3: When a set of experiences is phenomenally unified, there is a composite experience which subsumes them.

Premise 4: For any experience, there must be a subject. (Call this the ‘Ownership Principle’)

Premise 5: To undergo a composite experience involves undergoing all the experiences it subsumes.

6 (from 1 and 2): For all, or almost all, sets of things in our universe, the experiences belonging to its members are phenomenally unified.

7 (from 3 and 6): For all, or almost all, sets of things in our universe, there is a composite experience that subsumes all the experiences of its members.

8 (from 4 and 7): For all, or almost all, sets of things in our universe, there is a composite subject that is undergoing a composite experience which subsumes all the experiences of its members.

9 (from 5 and 8): For all, or almost all, sets of things in our universe, there is a composite subject that is undergoing all the experiences of its members.

I take premise 2 to be a plausible empirical observation, premise 4 to be a plausible principle about subjects (though not about *intelligent* subjects), and 3 and 5 to articulate the popular Bayne-Chalmers analysis of phenomenal unity. The upshot is that it is hard to accept the Micro-Unity Hypothesis without accepting free combination of subjects - though this might be just as good a *tollens* as a *ponens*.

An alternative thought would be that free combination of subjects derives somehow from the fact (if it is a fact) that, according to our best physics, fields are more basic than particles, or

the fact (if it is a fact) that our universe is really one, a single complex entity within which all divisions are superficial (see, e.g., Jaskolla and Buck 2012, Shani 2015, Goff 2017a, 209ff.; Nagasawa and Wager 2017; cf. Miller 2017; Albahari 2018; Shani and Kepler 2018).

Lots of panpsychists seem to like this line of thinking, but for my part I've always felt a bit sceptical, because I've yet to see anything like an actual derivation of, say, Experience Inheritance from any such doctrine - or, conversely, an actual derivation of *problem* for Experience Inheritance from the denial of such doctrines. Thus I am currently unconvinced by what we might call the 'cosmopsychism-only' position: that cosmopsychism solves the subject-summing problem in a way that micropsychism cannot.

My misgivings about the cosmopsychism-only position can perhaps be usefully expressed as a dilemma about what I'll call the 'Jamesian intuition'. This is the intuition that subjects (specifically subjects, not just objects in general) are metaphysically separate and independent things, which cannot be intelligibly combined, cannot share experiences, cannot have their experienced unified with each other's - the intuition that's pumped by the famous image of the 12 men thinking the 12-word sentence, uselessly crowding together. You might get to the cosmopsychism-only position by treating these intuitions as trustworthy at least when applied to fundamental subjects: if there were the conscious atoms that micropsychists posit, they'd be unable to combine with each other, so it's better to have only one fundamental subject, the cosmos, so that fundamental subjects don't *need* to combine. But here's the dilemma: do the Jamesian intuitions apply at all to things like us, mid-sized non-fundamental beings like horses and humans? If they do, and even human minds don't combine, can't have unified experiences with each other, can't share experiences with any larger or smaller subject, then it seems

impossible to intelligibly derive them from, or fit them into, the big cosmic subject. But if the Jamesian intuitions don't apply to beings like us, the position risks being both semantically and epistemically self-undermining. For surely beings like us are the defining paradigm instances of 'conscious subjects', so if some principles don't apply to things like us, but do apply to fundamental conscious universes or particles, then they're not principles about *subjects*. They might be interesting principles about something else - we might, in a Spinozistic mood, say that they apply to 'conscious substances', while clarifying that we are not ourselves substances. But even with such a change of terminology, there's something odd about the epistemic status of these principles too: where do they come from? We're well-acquainted with ourselves, but if we're not the things which these principles apply to, it's not clear how we would have trustworthy intuitive awareness of the principles in the first place. I'm not at all sure that the cosmopsychism-only position can't be made to work, but I can't see how to say anything much on this score with any confidence.

Setting aside cosmopsychism-vs.-micropsychism, I'm inclined to prefer a more *a priori* answer, that appeals to the nature of composition. The thought is that the reason consciousness is shared between parts and wholes, and indeed the reason other fundamental properties are shared between parts and wholes, is that the whole *is* the parts. Composition, on this view (see, e.g. Lewis 1991, pp.72–90, 1993, Baxter and Cotnoir 2014), is not a relation linking one set of things (the parts) to a further distinct thing (the whole): it is just our way of recognising that a single portion of reality remains the same reality whether we consider it as one or as divided into many parts.

On this view of composition, cases where a whole seems to lack the properties of its parts - like a composite of intelligent parts which is not itself intelligent, like a random group of human beings - are really cases where the property in question is defined in a division-relative way. The random group of humans *is* intelligent in a sense - it is intelligent on the human-being scale, since it instantiates the structure needed for intelligence across human-sized expanses of itself. It just isn't intelligent on the group scale, since it does not instantiate that structure across its whole expanse, and we put this by saying 'it' is not intelligent. It makes sense for intelligence to be division-relative in this way because its full analysis includes things like 'chooses the best action available in light of all available information', which means different things depending on which actions, and what information, counts as available - all the actions and information available to the group, or only some subset of them? But for truly fundamental properties no such analysis is available, so if they are instantiated by some part of the whole, they are automatically instantiated by the whole - and, equivalently, if they are instantiated in a local way by the whole, they are automatically instantiated by some part of it. The part-whole relationship here is neither holistic nor atomistic, by symmetrical.

That, at least, is the line of thought I like, though I do not think there is a huge difference between a form of panpsychism that grounds Experience Inheritance in composition-as-identity, and one which takes it to be among the basic natural facts. What is more important is to frankly accept that the subject-combination problem needs at least two solutions - one on which subjects combine freely and automatically, and one on which intelligent subjects combine only when the right structure arises.

Section 4: The Quality-Combination Problem

Next, let's consider what is usually called the quality-combination problem, or sometimes the 'palette problem'. It asks: given that there are only a few types of fundamental physical things, and given that the same fundamental physical things are involved in all my brain processes, how do my brain processes come to display such a wide variety of phenomenal qualities? My favourite statement of it is Lockwood's remark that "it is inconceivable that an artist, however skilled, should conjure the simulacrum of a Turner sunset from a palette containing only black and white paints." (1993, p.276)

It is useful, I think, to consider the palette problem together with another problem sometimes called the 'Revelation Problem'. This asks: given that there are trillions upon trillions of microphysical events going on in my brain, which according to the panpsychist are all associated with microexperiences of some sort, why do I not have introspective access to trillions upon trillions of experiential details? Why, in short, does the immense complexity of my brain state, which panpsychists say is reflected in an immensely complex experiential state, seem to be concealed from me?

I think the best approach for panpsychists is to treat these problems as two sides of the same coin - to say, in essence, that the diversity of phenomenal qualities *is* our way of introspectively apprehending the immense complexity of our brain, that the reason we do not seem to have access to trillions upon trillions of details is because those details blend together

into novel and diverse qualities. Indeed, ideas in this vicinity are fairly commonly discussed (see, e.g. Mørch 2018, pp.1079-1081): the issue is whether an account like this is defensible.

Let's spell out a little more clearly the idea that these two problems can, so to speak, solve each other. The idea is that under the right circumstances, two (or more) experiences may be experienced together by a subject, in such a way that they 1) are not introspectively distinguishable from each other, and 2) each contribute intelligibly to the quality of the whole experience. We may put condition 1) by saying that the two experiences are *confused* with one another; despite undergoing both, the subject cannot separate them, and so cannot tell that they are experiencing two distinct things. We may put condition 2) by saying that the two experiences are *blended* into the whole experience; the quality the subject is aware of incorporates the qualities of both.

Saying that the two experiences are 'experienced together' is here meant to simply convey phenomenal unity. When I see green to the left and blue to the right, I have both experiences together; they are unified into a composite experience. Of course in familiar cases like this, the subject having the composite experience is able to distinguish the two elements, to attend to them separately and be aware that they are experiencing two distinct things. The hypothesis we are considering for the panpsychist is that this is not an automatic or inevitable consequence of the experience being composite: that just as the blue and green might be experienced as side-by-side, or as above-and-below, or in any number of other arrangements, they might also be experienced without any such distinction ('in the same place'), i.e. as a single patch of turquoise. The claim is that this is also a way of experiencing blue and green together, even though they are confused with each other. Crucially, the fact that turquoise is a blend of

blue and green, i.e. that both blue-ness and green-ness are phenomenologically present *in* the experience of turquoise, allows us to say that the blue and the green are still being experienced: they have not vanished, replaced by a new quality, they are simply being experienced together in a specific way.

I should note that while colours are a good example to work with, as providing a sort of ‘proof of concept’, it may be that the actual brain structures responsible for colour-experience work in unexpected and surprising ways, and perhaps do not actually involve phenomenal blending. Indeed, even if they do involve phenomenal blending, there are likely a whole bunch of complications to the simple, vaguely painterly example just given. But the panpsychist’s job here is not to give an accurate story of how colour experience arises, from the electron upwards. It is to indicate a coherent *sort* of process, which could both produce qualitative diversity out of sparse ingredients, and also explain the absence, the seeming disappearance, of microstructural detail. I think the confusion/blending picture does that very neatly, and I think reflection on colours - and, perhaps even more so, on smells, tastes, and vague emotional moods - makes it very plausible that some process like this really does go on in our brains. In *Combining Minds* I express this as the following two hypotheses:

Radical confusion hypothesis (RCH): What makes human experience seem to be relatively coarse-grained is not an actual lack of fine-grained detail, but rather the fact that all of its component microexperiences are radically confused with one another. (2019, p.125)

Small palette hypothesis (SPH): All the phenomenal qualities experienced by humans and other beings arise from blending different combinations of the small range of basic phenomenal qualities experienced by microsubjects. (2019, p.125)

People have offered various *a priori* objections to both blending and confusion; I address these in *Combining Minds*, as well as in a couple of dedicated papers on the two notions (2014, Forthcoming). I don't think any of them succeed, and for the most part I think they're technical enough not to be worth the effort of setting up to respond to here. But one obvious question looming over the hypotheses I've sketched here is: when does blending/confusion occur? What are the 'right circumstances' alluded to above? I think the most plausible answer is that blending/confusion is the default for phenomenally unified experiences: it's only when a subject has the right kind of informational structure that elements within its experience can be distinguished. Our brains have a lot of the right kind of structure; plausibly animal brains do too. Without structure like that, our experience would just a vast buzzing roar, a huge blend within which every experience is submerged, rather like the experiential "white noise" that Peter Godfrey-Smith suggests might have been the evolutionary predecessor of sentience (2016, pp.95-97, cf. Ginsburg and Jablonka 2007). If that's right, then probably most macroscopic objects - tables, clouds, raindrops, and random mereological aggregates thereof - have that kind of experience: a single undifferentiated experience, whose quality is a blend of all the experiences going on within that whole. (Perhaps crystals, or bubbles, or biological cells, have their own unimaginable forms of phenomenal structure, too: I don't know.) The evolution of brains, and more broadly of nervous systems, would then have been not a process of

consciousness appearing from nothing, but a process of distinctions and contrasts slowly coalescing out of this kind of undifferentiated blend.

Section 5: The Structure-Combination Problem

Finally, I want to say a bit about the ‘structure-combination’, or ‘structural mismatch’, problem. I think this is the most open-ended problem, to the point that it’s almost not really a ‘problem’ and more just an ‘open question’. To the extent that it is a problem, it is standardly put as the idea that the physical structure of the brain somehow does not match the phenomenological structure of consciousness (see, e.g. Chalmers 2017, pp.190-191). Of course, most things have lots of sorts of structure, so it is not too surprising for the brain-mind to be structured in multiple ways. But there is at least a difference in which sorts of structure are highly salient from which perspectives: someone looking at a brain could hardly fail to notice its division into hemispheres, but we have no inkling of this introspectively. By contrast, a moment’s experience of a human mind impresses upon us the division among different sensory modalities, but this division is hard to discern from physically examining the brain. I don’t think there is really much basis here for an *objection* to panpsychism, but it is still a legitimate question whether panpsychists can offer a satisfying explanation of this kind of fact.

The beginning of an answer was already suggested toward the end of the last section: informational structure. That is, it seems independently plausible - indeed, almost obligatory, to explain the way our experience connects with our behaviour - that the way simple experiences build up to complex ones is heavily depend on the way that they carry information about each

other, the way that differences in each make a difference to the others, and in particular in the way that differences in one make different differences from the differences others make. In *Combining Minds* I expressed this as the ‘Informational Structure Hypothesis’:

Informational structure hypothesis (ISH): The overall structure manifest in human consciousness corresponds to the structure of information-processing in the human brain, not to its gross physical structure. (2019, p.125)

This is not remotely a new or surprising idea: panpsychists have been suggesting it in response to this kind of structure-combination question for years (see, e.g. Chalmers 1996, 284–292, 2017, 209–210; Gabora 2002). The challenge is to say more than this, to develop this sketchy idea both to make it more informative and to test whether it remains coherent and defensible.

Unfortunately, there’s a limit to how far panpsychists can really develop the idea of informational structure as determining conscious structure by themselves. The topic swiftly shades into the very broad topic ‘how is consciousness structured and why?’, and that is not a specifically panpsychist question, or indeed a specifically philosophical question. It’s a question for ongoing work across multiple disciplines, and it’s a question that arguably we are still very early in our investigations into. Famously, after all, it is the study of the world’s most complex known object, and I think the most that can be demanded of panpsychists is to be able to keep up with and go along with the slow piecemeal progression of this multi-disciplinary endeavour.

I try to do a bit of this keeping up with and going along with in *Combining Minds*. Partly it’s a matter of giving panpsychism-friendly treatments of phenomena where minds seem to be in various ways dissociated or integrated, like the split-brain syndrome, or collective agency, or

Dennett's multiple drafts model. Partly it's about trying to draw connections among seemingly disparate phenomena, like conscious unity and our awareness of other minds. And partly it's about the provision and criticism of metaphors. Let me provide an example of the latter.

Coleman says that "to take us toward the combination of the panpsychic ultimates . . . the metaphor-model I will appeal to is that of paint patches on a canvas" (2012, 157). I call this the 'canvas' model: each component experience gets "placed" at a certain spot in a pre-existing phenomenal space. My basic objection to this model is that when we paint on a canvas, a crucial role is played by the canvas itself—but where does this receptacle come from? Why does it have these particular locations available, this many of them, arranged in these dimensions? I worry that this kind of metaphor makes combination too easy: the real challenge is to see the phenomenal field itself, the space within which our experiences appear, as something constructed by the combination of our part's phenomenal fields. The paint patches must build the canvas.

My preferred metaphor appeals instead to the way that animators form images by superimposing multiple transparent layers (a metaphor also used by Lee 2019). Each subject, both parts and whole, has its own phenomenal field, within which different elements are laid out, some central and some peripheral. The whole's field is formed from all of the parts' fields superimposed upon one another: I call this the 'superimposition' model. Of course, it immediately raises the question of how these layers are aligned: with no pre-existing canvas to organise them, what decides how the internal structure of one field should map onto that of the others? Does the attentional focus of one component map onto the attentional periphery of another? Does the visual field of one component map onto the auditory experiences of another?

Unless something can fix the component fields into particular relations, the composite's field will be underdetermined.

I welcome this underdetermination, though. I think this provides a nice way to think about the confusion and blending discussed in the last section: the default outcome of superimposing multiple phenomenal fields, with nothing to fix their relative orientations, is a sort of inchoate blurr, in which each point in each field is equally and indiscriminately connected to each point in each other field - the 'experiential white noise' of inanimate object experience. What we have - structured consciousness, with phenomenal contrast among elements - comes from some sort of relation between the superimposed layers, that fixes their relative orientation by connecting specific elements in one field with specific elements in another. Metaphorically we might think of this as 'pinning' two layers together at a certain point, and this, I suggest, is what the right kind of informational relations do. Why this happens, I don't know - in part because we don't yet really know what kind of informational relations do this, or even how to think about information, and what we mean by 'informational relations'. It may be that the ultimate explanation will be *a priori*, once we understand phenomenology and information theory well enough. It may alternatively be that it all stems from some very basic but metaphysically contingent fact about causal powers in this universe - that it is just the nature of, say, the electromagnetic force that it effects a sort of phenomenal binding among microexperiences, which then ramifies upwards through into complex structures. My aim in this section is not so much to answer these questions as to make a case that panpsychists are well-placed to be active participants in the, likely centuries-long, process of answering them.

References

- Albahari, M. (2018). “Grounding the World in NonDual Awareness: A Metaphysical Model for Advaita Vedanta.” In W. Seager (ed.), *The Routledge Handbook of Panpsychism*. New York: Routledge.
- Bayne, T. (2010). *The Unity of Consciousness*. Oxford: Oxford University Press.
- Bayne, T., and Chalmers, D. (2003). “What Is the Unity of Consciousness?” in A. Cleeremans (ed.), *The Unity of Consciousness: Binding, Integration, Dissociation*. Oxford: Oxford University Press.
- Baxter, D., and Cotnoir, A. (eds.) (2014). *Composition as Identity*. Oxford: Oxford University Press.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. 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: 179-214.
- Coleman, S, and Alter, T. (2020). “Panpsychism and Russellian monism.” In W. Seager (ed.) *The Routledge Handbook to Panpsychism*, New York: Routledge: 230-242.
- Coleman, S. (2012). “Mental Chemistry: Combination for Panpsychists.” *Dialectica* 66 (1): 137–166.
- 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: 249-282.
- Dainton, B. (2000). *Stream of Consciousness: Unity and Continuity in Conscious Experience*. London: Routledge.
- Gabora, L. (2002). “Amplifying Phenomenal Information: Toward a Fundamental Theory of Consciousness.” *Journal of Consciousness Studies* 9 (8): 3–29.
- Ginsburg, S., and Jablonka, E. (2007). “The Transition to Experiencing: 1. Limited Learning and Limited Experiencing.” *Biological Theory* 2 (3): 218-30.
- Godfrey-Smith, P. (2016). *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*. New York: Macmillan.
- Goff, P. (2009). “Why Panpsychism Doesn’t Help Us Explain Consciousness.” *Dialectica* 63 (3): 289–311.

- 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*. New York: Springer.
- Goff, P. (2017-a). *Consciousness and Fundamental Reality*. Oxford: Oxford University Press.
- Goff, P. (2017-b). “The Phenomenal Bonding Solution to the Combination Problem.” In G. Bruntrop and L. Jaskolla (eds.), *Panpsychism: Contemporary Perspectives*. Oxford: Oxford University Press: 283-304
- Goff, P., and Roelofs, L. (Forthcoming). “In Defence of Phenomenal Sharing.” In J. Bugnon, D. O Conaill, and M. Nida-Rümelin (ed.) *The Phenomenology of Self-Awareness and Conscious Subjects*. Routledge.
- Howell, R. (2015). “The Russellian Monist's Problems with Mental Causation.” *Philosophical Quarterly* 65 (258):22-39.
- Jaskolla, L., and Buck, A. (2012). “Does Panexperientialistic Holism Solve the Combination Problem?” *Journal of Consciousness Studies* 19 (9–10): 190–199.
- Kammerer, F. (2015). “How a Materialist Can Deny That the United States is Probably Conscious – Response to Schwitzgebel.” *Philosophia* 43 (4): 1047-1057.
- Lee, A. (2019). “The Microstructure of Experience.” *Journal of the American Philosophical Association* 5 (3): 286-305.
- Lewis, D. (1976). “Survival and Identity.” In A. Rorty (ed.), *The Identities of Persons*. Berkeley: University of California Press.
- Lewis, D. (1991). *Parts of Classes*. Oxford: Basil 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, UK: Cambridge University Press.
- Locke, J. (1836/1689). *An Essay Concerning Human Understanding*. London: T. Tegg and Son.
- Lockwood, M. (1993). “The Grain Problem.” In H. Robinson (ed.), *Objections to Physicalism*. Oxford: Oxford University Press.
- McQueen, K. (2015). “Mass Additivity and A Priori Entailment.” *Synthese* 192 (5): 1373–1392.
- Mendelovici, A. (2020). “Panpsychism’s Combination Problem Is a Problem for Everyone.” In W. Seager (ed.), *Routledge Handbook of Panpsychism*. New York: Routledge: 303-316.
- Miller, G. (2017). “Can Subjects Be Proper Parts of Subjects? The De- combination Problem.” *Ratio* 30 (2): 1–18.

- Miller, G. (2018). “Forming a Positive Concept of the Phenomenal Bonding Relation for Constitutive Panpsychism.” *Dialectica* 71 (4): 541–562.
- Mørch, H. H. (2014). *Panpsychism and Causation: A New Argument and a Solution to the Combination Problem*. Doctoral Dissertation, University of Oslo.
- Mørch, H. H. (2018). “Is the Integrated Information Theory of Consciousness Compatible with Russellian Panpsychism?” *Erkenntnis* 84 (5):1065-1085.
- Mørch, H. H. (2019). “Is Consciousness Intrinsic?: A Problem for the Integrated Information Theory.” *Journal of Consciousness Studies* 26 (1-2):133-162.
- Nagasawa, Y., and Wager, K. (2017). “Panpsychism and Priority Cosmopsychism.” In G. Brüntrup and L. Jaskolla (eds.), *Panpsychism: Contemporary Perspectives*. Oxford: Oxford University Press.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Roelofs, L. (2014). “Phenomenal Blending and the Palette Problem.” *Thought: A Journal of Philosophy* 3:59–70.
- Roelofs, L. 2016. “The Unity of Consciousness, Within and Between Subjects.” *Philosophical Studies* 173 (12): 3199–3221.
- Roelofs, L. (2019). *Combining Minds*. New York: Oxford University Press.
- Roelofs, L. (2020). “Can We Sum Subjects? Evaluating Panpsychism’s Hard Problem.” In W. Seager (Ed.), *The Routledge Handbook of Panpsychism*: 245-258.
- Roelofs, L. (Forthcoming). “Consciousness, Revelation, and Confusion: are constitutive panpsychists hoist by their own petard?” *Dialectica*.
- Rovane, C. (1998). *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton, NJ: Princeton University Press.
- Seager, W. (2010). “Panpsychism, Aggregation and Combinatorial Infusion.” *Mind and Matter*, 8(2): 167-184
- Schechter, E. (2013). “Two Unities of Consciousness.” *European Journal of Philosophy* 21 (2): 197–218.
- Schechter, E. (2015). “The Subject in Neuropsychology: Individuating Minds in the Split-Brain Case.” *Mind and Language* 30 (5): 501–525.
- Schwitzgebel, E. (2015). “If Materialism Is True, the United States Is Probably Conscious.” *Philosophical Studies* 172: 1697–1721.

Shani, I. (2015). “Cosmopsychism: A Holistic Approach to the Metaphysics of Experience.” *Philosophical Papers* 44 (3): 389-437.

Shani, I., and Keppler, J. (2018). “Beyond Combination: How Cosmic Consciousness Grounds Ordinary Experience.” *Journal of the American Philosophical Association*.

Shoemaker, S. (1997). “Parfit on Identity.” In J. Dancy (ed.), *Reading Parfit*. Oxford: Blackwell.

Shoemaker, S. (2003). “Consciousness and Co-Consciousness.” In Axel Cleeremans (ed.), *The Unity of Consciousness: Binding, Integration, Dissociation*. Oxford: Oxford University Press.

Strawson, G. (2006). “Realistic Monism—Why Physicalism Entails Panpsychism.” *Journal of Consciousness Studies* 13 (10–11): 3–31.

Strawson, G. (2009). *Selves: An Essay in Revisionary Metaphysics*. Oxford: Oxford University Press.

Tononi, G. (2012). “Integrated Information Theory of Consciousness: An Updated Account.” *Archives Italiennes de Biologie* 150 (2–3): 56–90.

Tye, M. (2003). *Consciousness and Person: Unity and Identity*. Cambridge, MA: MIT Press.