How Exactly does Panpsychism Help Explain Consciousness?

Abstract: There has recently been a revival of interest in panpsychism as a theory of consciousness. The hope of the contemporary proponents of panpsychism is that the view enables us to integrate consciousness into our overall theory of reality in a way that avoids the deep difficulties that plague the more conventional options of physicalism on the one hand and dualism on the other. However, panpsychism comes in two forms – strong and weak emergentist – and there are arguments that seem to show that weak emergentist panpsychism faces problems analogous to those of physicalism whilst strong emergentist panpsychism faces problems analogous to those of dualism. In this paper, I will develop a new hybrid of the strong and weak emergentist forms of panpsychism, a view according to which subjects of experience are strongly emergent but their phenomenal properties are weakly emergent. I will argue that this hybrid view manages to avoid the challenges facing both physicalism and dualism, and the analogues of those challenges that seem to undermine standard forms of panpsychism.

There has recently been a revival of interest in panpsychism as a theory of consciousness. The hope of the contemporary proponents of panpsychism is that the view enables us to integrate consciousness into our overall theory of reality in a way that avoids the deep difficulties that plague the more conventional options of physicalism on the one hand and dualism on the other. However, panpsychism comes in two forms – strong and weak emergentist – and there are arguments that seem to show that weak emergentist

panpsychism faces problems analogous to those of physicalism whilst strong emergentist panpsychism face problems analogous to those of dualism. If this proves to be the case, it is hard to see how panpsychism advances the debate.

In this paper, I will develop a new hybrid of the strong and weak emergentist forms of panpsychism, a view according to which subjects of experience are strongly emergent but their phenomenal properties are weakly emergent. I will argue that this hybrid view manages to avoid both the challenges facing physicalism and dualism, and the analogues of those challenges that seem to undermine standard forms of panpsychism. The result is a theory of consciousness we should take very seriously indeed.

Section I gives background. Section II introduces 'hybrid cosmopsychism,' and argues that it avoids the problems that plague other forms of panpsychism. Section III takes a deep dive into the details of hybrid cosmopsychism. Section IV explains why the theory is a form of *cosmo*psychism rather than *micro*psychism (these terms will be explained below). Section V is a brief conclusion.

I

The word 'consciousness' is a little ambiguous. Throughout this paper, I will use 'consciousness' exclusively to mean *phenomenal consciousness*, states which are essentially characterized by what it's like to have them. Pleasure, pain, visual and auditory experiences are fairly uncontentious examples of phenomenally conscious states.

The problem of consciousness is the challenge of accounting for how (phenomenal) consciousness fits into our overall theory of reality. The reality of consciousness is hard to deny: nothing is more evident than the reality of one's own feelings and experiences. And so it seems that consciousness must fit into reality *somehow*; the challenge is to explain exactly how. The two traditional options are physicalism and dualism. Physicalists believe that the facts of consciousness can be accounted for in terms of the facts of physical science.[1] Dualists believe that conscious states are non-physical properties, residing either in the brain (property dualism) or in a non-physical individual (substance dualism). Both of these two traditional options face deep difficulties, which is what makes the problem of consciousness so hard.

The most discussed worry for dualism is the *causal exclusion problem* (Malcolm 1968, Kim 1989, Papineau 2000). Many philosophers believe that we have empirical reason to accept that the physical world is *causally closed*, that is to say that every physical event has a sufficient physical cause. If this is true, if everything I do has a sufficient physical cause (e.g. in terms of neurophysiological process in my brain), then it seems that putative non-physical consciousness has nothing left to do, no role to play in generating my behavior. The dualist seems to be driven either to *epiphenomenalism* (the view that consciousness has no causal impact on the physical world) or to *systematic over-determination* (all the effects of consciousness are systematically overdetermined, as every event caused by consciousness also has a sufficient physical cause). Many take these options to be intolerable.

[1] I do not mean by this that physicalists are committed to there being an a priori entailment from the facts of physical science to the consciousness facts; the popular phenomenal concept strategy (Loar 1990/1997, Balog 1999, Papineau 2002, Diaz-Leon 2008) would deny this. Rather I mean that, for the physicalist, the postulations made to account for the data of physical science are also sufficient to account for consciousness. I give a more detailed definition of physicalism in Goff 2017a: Ch. 2. See next footnote for further clarification.

The problem with physicalism is that there seems to be an explanatory gap between the facts described by physical science and the facts of consciousness. The charge is not merely that our current theories are not up to the task, but that there is an in principle bar to the latter being explained in terms of the former, rooted in the very different kinds of concepts we use to characterize physical processes on the one hand and conscious experiences on the other hand. The former concepts are quantitative and third-personal, whereas the latter are qualitative and first-personal. Proponents of the knowledge argument and the conceivability argument argue from this explanatory gap to the conclusion that the postulations of physical science alone are not enough to ground the facts of consciousness.[2]

Of course, in both of these cases, there is much controversy, with dualists and physicalists claiming that there are satisfactory responses to these arguments.[3] However, for philosophers who find these arguments compelling, there is strong motivation to look for an alternative theory of consciousness.

[2] The 'postulations' of physical science may refer to the dispositional properties expressed by the predicates of physics, or it may refer to categorical properties underlying those dispositions. As I define physicalism, a physicalist may postulate categorical grounds that in some sense take us beyond what physics reveals to us (to this extent, physicalism may resemble Russellian panpsychism, discussed below). But, unlike Russellian panpsychism, physicalists do not do this specifically to account for consciousness but only because they think that in general dispositions need categorical grounds. Hence, the categorical nature of physical properties will not have some special character tailored to account for consciousness. As I have defined physicalism previously (Goff 2017a: Ch. 2), this comes out as the thesis that physicalists do not commit either to phenomenal or to proto-phenomenal properties at the fundamental level (where proto-phenomenal properties are defined as being involved in facts that a priori entail facts about consciousness, where that entailment isn't wholly dependent on structural features of the grounding fact (where structural features are properties whose essential nature can be captured in a purely mathematico-causal vocabulary)) .

[3] On the physicalist side, the most popular response is the phenomenal concept strategy Loar 1990/1997, Balog 1999, Papineau 2002, Diaz-Leon 2008. For some dualist responses, see Chalmers 1996, Gibb 2015, Robinson 2019.

Against this background, the Russellian panpsychist steps in with a promise to avoid both the causal exclusion problem faced by dualism and the explanatory gap problem faced by physicalism. The view is so-called because it is inspired by certain claims Bertrand Russell made in *The Analysis of Matter* of 1927, although the view Russell defended here was not quite a form of panpsychism. Perhaps the best way to introduce Russellian panpsychism is to say that it has a negative component and a positive component. Let us take each of these in turn.

The negative component is the claim that physics tells us less than we might have thought about the nature of physical reality. Physical science identifies the causal roles associated with fundamental physical properties, such as mass, spin and charge, but doesn't tell us the essential nature of the properties that realise those causal roles. Mass, for example, is characterized in terms of gravitational attraction and resistance to acceleration, and charge in terms of attraction and repulsion. Physics tells us what mass and charge *do* – the causal roles they realise – but not what they *are*.

This negative aspect of Russellian panpsychism identifies a huge hole in our standard scientific story of reality. The positive proposal of Russellian panpsychism is to put consciousness in this hole: physical properties are, in their essential nature, forms of consciousness. Thus, Russellian panpsychism is a radically *non-dualistic* form of panpsychism. It is not the view that matter has physical properties (mass, spin, charge) on the one hand and experiential properties on the other. Rather the claim is that physical

properties like mass, spin and charge *are* forms of consciousness.[4] Physical science tells us what mass does, but, in terms of its essential nature, mass is a form of experience.[5]

Russellian panpsychism assumes a distinction between the causal role a property plays and its essential nature: what a property does versus what it is. Some reject that distinction. Couldn't it be that the essential nature of a property is *given* by its causal role? An affirmative answer to this question is made by proponents of *pan-dispositionalism*.[6] On this view, once you know everything this is to know about the causal role of mass, you know everything there is to know about what mass is.

Some Russellian panpsychists doubt the coherence of pan-dispositionalism. There is a line of argument, going back to Russell himself, which presses that on a pan-dispositionalist view everything is defined in terms of everything else, which, it is alleged, leads to a kind of vicious circularity.[7] But even if this argument fails and pan-dispositionalism is a coherent metaphysical option, this is consistent with Russellian panpsychism also being a coherent

[4] Alternately, the Russellian panpsychist may hold that physical property terms refer to dispositional properties, and hence that physical properties are realized by, rather than identical with, forms of consciousness. The disagreement between this view and the view described in the main text is not one of substance but rather regards how terms in physics are defined. I suspect it is indeterminate whether the linguistic practice of physical scientists is such that 'mass' refers to a dispositional property or to a categorical property in terms of the dispositions it realizes.

[5] Strawson 2006, Chalmers 2015, Brüntrup & Jaskolla 2016, Goff 2017a, Mørch 2019, Roelofs 2019, Seager 2019. See Goff 2019a for an accessible, non-academic book defending this view. A closely related view is Russellian *panprotopsychism*, which hopes to explain consciousness by postulating proto-phenomenal properties at the fundamental level (see footnote 2 for a definition of proto-phenomenal): Pereboom 2011, Coleman 2016, McClelland 2013. Russellian panpsychism and panprotopsychism are known collectively as 'Russellian monism'; see Goff & Coleman 2020 for an overview, and Alter & Nagasawa 2015 for a collection of essays on Russellian monism.

[6] Ellis 2001, 2002, Molnar 2003, Bird 2007, Mumford 2004.

[7] Russell 1927, Campbell 1976, Robinson 1982, Lowe 2006, Goff 2017a: Ch. 6. For an accessible non-academic version of this argument, see Goff 2017d.

metaphysical option. The crucial question is whether the theoretical attractions of the latter give us reasons to embrace it.

What are the theoretical attractions of Russellian panpsychism? The ultimate goal, of course, is to account for human and animal consciousness in terms of more basic forms of consciousness. The hope is that by doing this, we can avoid the problems that beset physicalism and dualism. We avoid the problems of physicalism because the arguments that press the explanatory gap target physical-science based accounts of consciousness, i.e. accounts which aim to explain consciousness in terms of the postulations made to account for the data of physical science, whereas Russellian panpsychists are instead trying to account for consciousness in terms of the 'hidden' essential nature of the physical world.[8] And we avoid the problems of dualism because consciousness is incorporated into the causally closed physical system; it is only once we distinguish physical processes from consciousness processes that causal exclusion problems arise. At least, this is the hope of Russellian panpsychists.[9] Unfortunately, it's not clear these problems can be dispensed with so easily.

As we have seen, Russellian panpsychists hope to account for human and animal consciousness in terms of more basic forms of consciousness. But how exactly is this done?

[8] Broadly speaking, physicalist accounts of consciousness fit into two categories, which David Chalmers (2002) dubbed *type-A* and *type-B*. Type-B physicalists hold that we do not need to close the explanatory gap to explain consciousness. Type-A physicalists do try to close the explanatory gap, but by attempting to explain consciousness in terms of causal roles. The distinctive approach of the Russellian monist is to try to close the explanatory gap by postulating a special essential nature to the properties underlying the causal roles identified by physical science. See footnotes 1 and 2 for further clarification of how I understand physicalism in contrast to Russellian panpsychism.

[9] See, for example, Chalmers 2015 and Goff 2017a. For an accessible, non-academic introduction to this argument, see Goff 2016.

There are broadly speaking two options. One option is to postulate basic laws of nature that bridge the gap between consciousness at the fundamental level and the consciousness of an animal. Thus, it might be simply a basic law of nature that when you have conscious particles arranged in such and such a way, consciousness associated with the whole system emerges. This is the *strong emergentist* option. The *weak emergentist* panpsychist, in contrast, tries to account for systems-level consciousness without appeal to such extra laws of nature. On this version of Russellian panpsychism, facts about human or animal consciousness are wholly constituted by facts about consciousness (and perhaps physical structure) at the fundamental level; the latter are nothing over and above the former, in something like the way that on physicalism the facts about consciousness are nothing over and above the facts of physical science.[10]

The problem is that there is reason to think that the strong emergentist panpsychist faces the causal exclusion problem endured by dualism, whilst the weak emergentist panpsychist faces the explanatory gap problem suffered by physicalism. If this proves to be the case, it seems that we've got nowhere.

These concerns are standardly pressed against the background assumption that the fundamental facts for the panpsychist concern fundamental particles bearing very simple forms of consciousness. We can call panpsychism so understood 'micropsychism.'[11] The

[10] Chalmers (2015) has a similar distinction between constitutive and non-constitutive forms of Russellian panpsychism. However, I want here to focus on whether or not extra laws are needed to account for emergent consciousness facts, and, whilst non-constitutive panpsychists will tend to postulate extra laws, this is not part of the definition of the view. Chalmers (2006) gives an account of strong and weak emergence in epistemological terms.

[11] Strawson 2006 gives a slightly different definition of 'micropsychism.'

challenge for the micropsychist is to bridge the gaps between particle-level consciousness and systems-level consciousness. We will later reject micropsychism but can work with it for the moment.

Why think strong emergentist panpsychism faces causal exclusion worries? If we suppose that the micro-level is causally closed, then the new systems-level consciousness that strongly emerges would seem to have nothing left to do in generating behavior, and as a result would seem to be rendered epiphenomenal (Chalmers 2015, Goff 2017a, b, 2019b).[12] One might also worry that once one commits to explaining human consciousness in terms of special laws of nature, one loses the motivation for adopting panpsychism. Why not just be a property dualist, bridging the gap between physical-science properties and biological consciousness via special laws of nature, rather than postulating consciousness everywhere?

Why think the weak emergentist panpsychist faces explanatory gap worries? One way of pressing this (Goff 2009, Chalmers 2016) is in terms of a variant of the zombie conceivability argument against physicalism. Whereas standard zombies are physical duplicates of humans or animals which lack consciousness altogether, *micro-experiential zombies* are physical

[12] By saying that the 'micro-level is causally closed,' I mean that every event either has a sufficient micro-level cause or has a sufficient cause that is wholly grounded in a micro-level event. One might also worry (thanks to Hedda Hassel Mørch for raising this concern) that if all the causal structure of the physical is at the micro-level, then there is no macro-level physical causal structure for the strongly emergent conscious states to realise, and hence the strongly emergent physical states will count as non-physical rather than physical (given that 'physical' states, for the Russellian panpsychist, are the states that realise the causal structure discerned by physical science). However, even if micro-level causal closure is true, there will still be macro-level causal structures realised by the micro-level, and it will be coherent to suppose that those macro-level causal structures are also realised by strongly emergent consciousness (in which case there will be a kind of over-determination of these macro-level structures by two distinct realisation bases). Even if strongly emergent states are epiphenomenal, and hence do not realise any physical causal structure, we could say that they are physical in virtue of being of the same essential nature as the physical states from which they strongly emerge.

duplicates of humans such that (A) all of their most basic parts have conscious experience, but (B) there is no systems-level consciousness, i.e. no consciousness associated with any macro-level part of the organism. Panpsychist zombies seem prima facie just as conceivable as regular zombies. If the possibility of regular zombies follows from their conceivability, then the same would seem to be true of micro-experiential zombies; and if the possibility of regular zombies is inconsistent with the truth of physicalism, then the possibility of micro-experiential zombies would seem to be inconsistent with the truth of weak emergentist panpsychism.

This is, of course, a particularly worrying problem given that panpsychism is often motivated via an employment of the zombie argument (to reject physicalism). If that very argument, in a slightly modified form, rules out panpsychism, we seem to have made no progress. Furthermore, there are arguments which purport to show that *any* reductive account of a conscious subject, even one in terms of more basic forms of consciousness, must necessarily fail (Goff 2015, 2016, 2019d, Nida-Rümelin 2014). This is one form of 'the combination problem', the banner for a broad range of challenges to the panpsychist's attempt to bridge the gap from micro-level consciousness facts to the familiar facts of human and animal consciousness.[13]

These are certainly very serious challenges. However, I am not entirely persuaded that the problems they raise are *as* serious as the corresponding challenges facing dualism and physicalism. As Hedda Hassel Mørch has pointed out (2014: 3.1, 2023: 4.4.3), it is not so

[13] The term 'combination problem' is from Seager 1995. For more detail on the combination problem, see Chalmers 2016 and Goff 2017a: chs. 7-8, Goff et al 2017/2022.

clear that there is a strong empirical case for *micro*physical causal closure, as opposed to a more general thesis of physical causal closure. This is difficult to assess, as the case for causal closure is often stated but rarely defended. But one defence of causal closure (McLaughlin 1994: 278) is the 'no-gap' argument: an inductive argument which starts from the putative observation that we never find gaps in the physical causation observed in the brain. It is open to the emergentist panpsychist to hold that human conscious states *are* physical states of the brain, just ones whose existence is not wholly grounded in facts about their parts (the whole is more than the sum of its parts). It's not immediately obvious why the no-gap argument would rule out strongly emergent physical states.[14]

Regarding the motivation for adopting strong emergentist panpsychism over dualism, I would suggest that whilst panpsychism seems at first to be a rather extravagant thesis, upon further reflection it turns out to be much simpler and more elegant and unified a picture of reality than that offered by the dualist.[15] On a dualist view, there is a radical division between two fundamentally different kinds of property; on the panpsychist view, all fundamental properties are of the same kind: experiential properties. Even in its strong emergentist form, therefore, we may have reason to prefer panpsychism to dualism. (See Mørch 2014, Ch. 3, 2023: 4.4.3 for further arguments that strong emergentist panpsychism is more plausible than dualism.)

[14] It is plausible that neurophysiological properties are essentially defined as complex properties composed of micro-physical properties. But this is consistent with the fusion form of strong emergentism discussed in footnote 34 (at least if 'X is composed of Y' indicates merely a part/whole relationship and doesn't entail that Y is more fundamental than X).

[15] Goff 2017a: Ch. 7. For an accessible, non-academic version of this argument, see Goff 2017c.

Turning to weak emergentist panpsychism, the putative gap between particle-level and systems-level consciousness is arguably less severe than the gap between the facts of physical science and the facts of consciousness. In the case of the physical/consciousness gap, the kinds of concepts employed on either side of the gap are very different: third-personal, quantitative concepts on the one side, first-personal, qualitative concepts on the other. In the case of the particle-consciousness/systems-level-consciousness gap, the same kinds of concepts are employed on each side: first-personal, qualitative concepts. It is also noteworthy that, whilst a zombie argument seems to apply to both physicalism and weak emergentist panpsychism, it is much less obvious that a version of the knowledge argument applies to weak emergentist panpsychism.[16]

What I want to consider for the rest of the paper, however, is what options are available if one is persuaded that both strong and weak emergentist panpsychism, at least in their standard forms, fail due to the problems discussed above. I believe that there is a form of panpsychism that avoids these concerns, and it is to this that we now turn.

II

In the last section, we worked with the assumption that fundamental entities exist at the micro-level. However, for reasons I will explain in section IV, the panpsychist view I want to defend here is a form of *cosmopsychism*.[17] Contemporary cosmopsychist views build on the

[16] Goff 2019b.
[17] Examples of cosmopsychism include Mathews 2011, Shani 2015, Nagasawa & Wager 2016, Goff 2017a: Ch. 9, 2019d, Shani & Kepler 2018.

priority monism developed by Jonathan Schaffer (2010). Philosophers have often assumed that fundamental entities exist at the micro-level, such that all facts are grounded in facts about arrangements of micro-level entities. According to priority monism, however, there is just one fundamental entity: the universe as a whole. All facts are grounded in facts about the universe. There are a variety of ways in which we this could be spelt out, but I will construe priority monism in terms of a field ontology, according to which fundamental reality is made up of universe-wide fields, and particles are identified with local excitations of these fields. On a priority monist view, we can hold that these fundamental fields are basic attributes of the one fundamental individual: the universe.

Of course, priority monism does not entail panpsychism, but there is a closely related form of panpsychism: cosmopsychism, the view that the universe is conscious and that all facts depend on facts about the conscious universe.[18] On the micropsychist version of Russellian panpsychism, very simple forms of consciousness are the essential nature of the physical properties of particles. On the cosmopsychist view I will explore here, very complex forms of consciousness are the essential nature of the universe-wide fields born by the cosmos. Imagine the complete description of reality in the terms of fundamental physics: an incredibly complicated story of patterns of excitation in fundamental fields. That very complicated structure, on the view under consideration, is realised by the experience of the universe.

[18] Forms of cosmopsychism according to which all facts are grounded in facts about the universe-subject are forms of priority monism. However, cosmopsychists who think that local (i.e., non-cosmic) subjects *strongly* emerge from the universe may deny that local subjects are *grounded* in the universe, in which case they wouldn't count as priority monists (as Schaffer defines priority monism). Schaffer (2017) himself thinks grounding relations are underwritten by basic laws, and hence holds both that local subjects are grounded in the universe and that local subjects strongly emerge (as I am defining strong emergence). However, many others, myself included, would take the strong emergence of local subjects to entail their fundamentality.

It might be worth emphasising that this is not – or at least need not be – pantheism or the postulation of a Hegelian world-soul. In other work, I have developed a form of cosmopsychism on which the universe is goal-directed, a view I motivate in terms of the need to explain the fine-tuning of physics for life.[19] I call this view 'teleological cosmopsychism.' However, if one is just trying to explain ordinary biological consciousness, as we are here, it is unlikely that one will end up ascribing intelligence or agency to the cosmos. When engaging with panpsychism, we should be careful not to model all consciousness on the highly unusual, because highly evolved, consciousness of human beings.

Strictly speaking, cosmopsychism avoids the combination problem discussed in the last section. If we're not trying to get from conscious particles to systems-level consciousness, then we don't have to worry about the putative explanatory gap that holds between these two levels. But a moment's further reflection makes it clear that we've merely pushed the lump to another part of the carpet: we now face an explanatory gap between the consciousness of the universe and the consciousness of humans and animals. It seems perfectly conceivable that we might have a conscious universe, with experience corresponding to the structure of basic physics, without any of the *parts* of the universe being conscious. Perhaps in some sense the universe would instantiate human experience, or at least experience corresponding to the physical structure of human bodies and brains (more on this soon). But what we surely want to make sense of are *multiple subjects*

[19] Goff 2018 is an accessible, non-academic version of the basic case for this; Goff 2019c is my first academic article outlining the view. Goff forthcoming is my most developed exploration of the view, and is aimed at both an academic and a popular audience.

corresponding to different people and animals. According to our pre-theoretical understanding of things, there are at least seven and a half billion conscious subjects in the world, corresponding to the 7.5 billion people in the world. A mere commitment to a conscious universe seems to give us only one. Whereas micropsychism faces a 'combination' problem, cosmopsychism faces this 'de-combination' problem (Chalmers 2016, Goff 2017a, Albahari 2019).[20]

As we found in the case of micropsychism, this new explanatory gap is only a problem for a weak emergentist form of cosmopsychism. We could instead adopt a strong emergentist form, according to which there are basic laws of nature which ensure that when the conscious universe is in certain specific states, *new* forms of consciousness emerge corresponding to certain of the universe's parts. But – here we go again! – it seems we would then be back to worries about causal closure: if the level of basic physics is causally closed, there is no causal work left for these new forms of consciousness to do.

The way forward I want to explore is a hybrid of the strong and weak emergentist approaches: strong emergentism about subjects combined with weak emergentism about the conscious states of emergent subjects. We can call this view 'hybrid cosmopsychism.'

Let's begin with by clarifying terminology and initial assumptions. I take *phenomenal properties* to be ways of experiencing, distinguished by what it's like to have them. *Conscious subjects* are the bearers of phenomenal properties: a conscious subject is a thing

[20] The phrase 'de-combination problem' was coined by Miri Albahari in the paper referenced here, which was circulated a long time before publication.

such that there's something that it's like to be that thing. I will understand an *experience* to be a particular instance of a conscious subject bearing phenomenal properties. For the most part, I will talk as though there is an ontological distinction between a subject and its phenomenal properties, but, as we'll see below, this is not an essential commitment of hybrid cosmopsychism.

According to hybrid cosmopsychism, there are basic laws ensuring that, in certain conditions, new conscious subjects – new phenomenal property bearers – emerge from the universe. But these strongly emergent subjects do not appear with their own phenomenal property instances; rather they 'inherit' phenomenal property instances which previously belonged to the universe. That is to say, there are phenomenal properties, P_1, P_2…P_n, and an emergent subject E, such that at time T1, P_1, P_2…P_n belong to the universe and at a later time T2, P_1, P_2…P_n cease to belong to the universe and instead belong to E. As E persists through time, it continues to possess a small 'bubble' of the phenomenal properties of the fundamental fields, constantly gaining some from/losing some back to the universe around the edges.[21] At the moment E ceases to be a conscious entity in its own right – perhaps at the death of the organism – it relinquishes its phenomenal properties back to the universe. In this way, although there are strongly emergent *subjects* there are no strongly emergent phenomenal *properties*; rather phenomenal properties of the fundamental fields are

[21] I have framed the view in non-relativistic terms for the sake of ease of illustration. We might ultimately want to frame the laws discussed below in terms of spatiotemporal locations of emergent subjects. I'm grateful to Bradford Saad for raising this issue.

transferred from the universe subject to emergent subjects, before being relinquished back to the universe.[22]

The crucial advantage of hybrid cosmopsychism is that it accounts for the fact that systems with emergent subjects behave according to the same basic laws of nature as systems without emergent subjects. The fundamental properties driving the evolution of physical reality are the properties of the fundamental fields; their causal powers are tracked by fundamental physics, and they are unchanged by the presence or absence of emergent subjects. All that changes when a new subject emerges is that some properties once borne by one subject come to be borne by another subject. But given that the new subject is just bearing properties that would have been borne by the old subject if emergence hadn't occurred, and is bearing them in the exact same location that they would have been borne if emergence hadn't occurred, there are no grounds for thinking that the evolution of physical reality will be affected by the emergence of local, i.e., non-cosmic, conscious subjects. [23]

In one stroke, this removes any of the empirical difficulties associated with dualism and strong emergentist panpsychism. The experiences of strongly emergent subjects are not epiphenomenal, and nor is their causal work over-determined at the level of basic physics (the cosmic level, on a cosmopsychist view). Rather the universe shares its causal work with the strongly emergent subjects, thus avoiding causal exclusion. On dualism and standard

[22] Hybrid cosmopsychism makes the common, although not universal (Roelofs 2019), assumption that not all macro-level systems are conscious. In the absence of a compelling case for dropping this assumption, the benefits of hybrid cosmopsychism, in my view, justify making this assumption.

[23] Of course, the fact that some systems do and some systems do not involve emergent subjects is itself a big difference between them, and this difference is due to the laws of nature. My claim is simply that, on the view I'm defending, it is not surprising that the externally observable behaviour of systems can be predicted with the same principles regardless of whether or not they involve emergent subjects.

forms of strong emergentism, one would expect that such a radical change in nature as the strong emergence of radically new forms of macro-level consciousness would show up in physical systems associated with that emergence. We would expect physical systems that involve macro-level consciousness to operate according to new, or at least modified, laws of nature. This is not what we seem to find, and thus dualism and strong emergentist panpsychism would seem to be disconfirmed. In contrast, the expectation engendered by hybrid cosmopsychism matches what we in fact observe: systems with macro-level consciousness behave according to exactly the same laws of nature as systems that lack macro-level consciousness.[24]

At the same time, the strong emergentist element of hybrid cosmopsychism also removes the worries that plague weak emergentist forms of panpsychism. Let us refer to any subject non-identical with the cosmos as a 'local subject.' Although it is conceivable that that a conscious universe might exist in the absence of local subjects, it is *not* conceivable that:

- There is a conscious universe and basic laws of nature determining that, in actually instantiated conditions, local subjects emerge and inherit certain phenomenal properties of the universe.
- There are no local subjects.

The extra basic laws committed to by the hybrid cosmopsychist bridge the explanatory gap between cosmic consciousness facts and local consciousness facts.

[24] Is this solution not ad hoc? Haven't we just designed the theory such that a radical change in nature goes unobserved? This change is only unobservable *from the third-person perspective*. It is not at all surprising that a partial description of reality should leave some facts unknown.

In summary, the strong emergentist element allows hybrid cosmopsychism to avoids explanatory gap worries, whilst the weak emergentist element allows it to avoid causal closure worries. This is exactly the result we want.

III

In this section, we will explore the details of hybrid cosmopsychism.

A crucial component of the view is *qualia transference*. In a case of qualia transference, phenomenal property instances are transferred from one subject of experience to another. To take a toy example, we might imagine a wicked witch who regularly has headaches. They don't bother her, however, as she immediately transfers the horrible phenomenal properties involved to a hapless passerby, who suffers the pain in her place. As I am imagining the story, it is not simply that the witch's headache disappears to be replaced by a qualitatively indiscernible headache in the passerby. Rather, the phenomenal property that leaves the witche's consciousness is *numerically identical* with the phenomenal property that appears in the consciousness of the passerby.[25]

[25] I think this position is compatible both the view that properties are universals and with the view that properties are tropes. On the former view, so long as the phenomenal property of the witch pre-transfer – call it 'W' – is qualitatively identical to that of the passerby post-transfer – call it 'P' – then it follows that the W is numerically identical to P. It is slightly non-standard to combine tropes with substance-attribute theory, but not unheard of (Martin 1980; Heil 2003; Lowe 2006); note we will explore a bundle version of the view below.

The witch analogy is grossly over-simplified, just to give the basic idea.[26] Let us spell out in more detail how hybrid cosmopsychism makes use of qualia transference. Contemporary neuroscience suggests that the structure of human consciousness corresponds to high-level information structures in the brain. The global workspace theory (Baars 2002), for example, holds that consciousness corresponds to information that's 'broadcast' throughout the brain, i.e., that's broadly available for many different systems in the brain. While they are in some sense realized by electro-chemical processes, these information structures abstract away from neurophysiological details, i.e., from facts about calcium chambers, chemical composition of neurotransmitters, etc.[27]

The structures of human consciousness, therefore, are not the structures of basic physics. But plausibly these structures are in some sense present in the structure of physics. Consider the complete story of the universe in the language of basic physics. Focus on the bit of that structure that's located in my head, and abstract away from a lot of micro-level detail leaving only coarse-grained causal structure. If you did this in the right way, you'd be able to find information structures isomorphic with my conscious experience. Given this, cosmopsychism implies that the experience of the universe-subject contains structures isomorphic with the structure of animal experience. In other words, if you take the rich and complex experience of the universe – which *ex hypothesi* underlies the structure of physics – focus on the bit of that experience that's located in a particular human head, abstract

[26] For one thing, the witch analogy seems to be assuming property dualism, unless she's magically transferring a bit of her physical brain to the victim.

[27] It is compatible with cosmopsychism that higher-level information structures are realised by more fine-grained physical states, so long as the more fine-grained physical states are states of, or grounded in states of, the universe.

away from a lot of micro-level detail, you'll be able to find structure isomorphic with the structure of that human's experience.

Of course, just because the structures of animal experience are in some sense present in the structure of cosmic experience, it doesn't follow that there are multiple subjects corresponding to each such structure. What we need to do if we want to account for the facts of animal consciousness is to formulate laws that determine that there are local subjects that inherit those aspects of cosmic experience that are structurally isomorphic with animal experience. How could this be done?

I suggest two principles specifying the kind of laws we need: the *Localization Principle* and the *Thinning Principle*. The Localization Principle is fairly straightforward: it says that we need a law determining that local subjects exist and only exist when certain precise conditions obtain. Which conditions? This is an empirical question, to be settled by our best theory concerning the physical correlates of local consciousness. The Thinning Principle is a little more nuanced; it says that we need a law ensuring that local subjects inherit a 'thinned-out' version of the experience contained in the spatial region they occupy, such that emergent subjects inherit only those aspects of experience that realise the right information structures (where the 'right information structures' are the ones introspection and neuroscience tell us correspond to human and animal experience). This notion of 'thinning out' is a conceptual innovation that we now need to clarify.

In a sense, the simplest case of qualia transference would be one in which all the phenomenal properties of one individual are transferred wholesale to another individual,

such that what it's like to be the first individual at T1 is exactly the same as what it's like to be the second individual at T2. But we might also imagine that *some aspects* of the total phenomenal properties of one subject could be transferred to another subject, whilst the other aspects remain with the original subject.[28] Consider the following case. Subject A has a detailed visual experience as of a lake surrounded by trees, with birds flying overhead. Some aspects of A's phenomenal properties are then transferred to B. The result is that both subjects end up with a 'thinned-out' version of the original experience; perhaps subject A ends up with an experience of a lake without trees and birds whilst subject B ends up with an experience of trees and birds without a lake.

We can now apply this to the theory under consideration. According to hybrid cosmopsychism, the fields of physics are incredibly complex universe-wide phenomenal properties, which are initially borne by the universe. As discussed above, it is plausible that structures isomorphic with my experience are to be found in the experience borne by the universe in the region of space where my brain is, if you abstract away from enough detail. We can suppose, then, that the phenomenal properties borne by *me* are identical with certain aspects of the phenomenal properties initially borne by the universe: those aspects that would remain if you started with the very busy experience corresponding to the total physical structure of the brain and then stripped out – in a very selective way – a great deal of detail. The purpose of the thinning law is to ensure that precisely the right aspects – those corresponding to the information structures that contemporary neuroscience tells us

[28] I'm not thinking of aspects as being a different category of thing from phenomenal properties. Rather some phenomenal properties are aspects of other phenomenal properties, e.g. the phenomenal property of phenomenal hue may be an aspect of the phenomenal property of phenomenal red.

correspond to my experience – cease to be borne by the universe and are instead borne by me.

This might initially seem a bit hard to make sense of, as though chunks of the universe are being taken away from it, which might give the impression that this ought to leave holes in the universe. But the idea is that the basic universe-wide phenomenal properties which we are identifying with the fundamental universe-wide fields of physics remain unaffected by qualia transference. All that changes is that certain aspects of those properties transfer from being born by one subject (the cosmos) to being borne by another (a local subject). Essentially, post-transference, the cosmos and the local subjects share the work of bearing the fields of physics.

Some (Albahari 2019) have worried that if cosmopsychism implies that the universe bears all of the conscious states of every local subject, this would imply that the universe has all of our thoughts, which would seem to imply that the universe has contradictory thoughts (given that people's views contradict). One advantage of hybrid cosmopsychism is that it avoids this concern, as the universe does not bear human consciousness. Rather, the universe bears those aspects of the phenomenal properties in our heads which are not borne by us. If the cognitive states of humans are grounded in their consciousness (Mendelovici 2018) – itself a controversial view – this is presumably at least in part because human consciousness embodies sophisticated information structures, structures which according to hybrid cosmopsychism emerge when the meaningless experience borne by the universe is thinned out in a highly selective manner. The kind of delicately-thinned-out-consciousness borne by human beings is never borne by the universe.

Isn't this dualism? It's true that the emergence of local subjects involves new entities coming into bear, but and all of the properties borne by these new entities are aspects of the intrinsic natures of the field of physics.[29] It seems to me appropriate, therefore, to interpret the theory as one on which a local subject is a physical object, occupying the same location as the phenomenal properties borne by that subject. In any case, what is important is not whether we call it dualism, but whether it avoids the problems traditionally associated with dualism.

But does hybrid cosmopsychism really avoid those worries? One might worry that this view implies epiphenomenalism because it's really the universe-wide phenomenal properties that are running the show, whilst the various subjects – whether the universe or local subjects – are merely passively taking in the show.

There are different ways to respond to this worry, depending on how you think about the relationship between objects and properties. Some philosophers think a property cannot exist by itself in the absence of some entity that *has* the property, in which case phenomenal properties only exist in so far as there are conscious subjects *experiencing* those properties. On this view, the objection collapses, as phenomenal properties cannot exist or cause anything independently of subjects.

[29] See O'Connor 2018 for a good discussion of the coherence of strong emergentism, including the coming to be of new substrata.

Other philosophers believe that objects are simply bundles of properties, and hence that a conscious subject is just a bundle of phenomenal properties. We could combine this kind of view with hybrid cosmopsychism by holding that at the fundamental level there are only phenomenal properties and facts about which aspects of those properties are co-experienced. On this hypothesis, the emergence of the first local subject results from phenomenal properties that were previously co-experienced with all phenomenal properties in existence coming to be co-experienced only with each other. Thus, we have a situation in which:

- At T1, phenomenal properties P1, P2…Pn are co-experienced with every other phenomenal property in existence
- At T2, P1, P2…Pn are co-experienced only with each other.

In this way, we can think of subjects as non-fundamental entities, derived from fundamental facts about co-experiencing. Despite this, this is still a form of strong emergentism about local subjects, given the reliance on fundamental laws to bring local subjects into being. For the rest of the paper I will assume a version of hybrid cosmopsychism according to which subjects and not reducible to bundles of properties, without meaning to imply that this view has any advantages over the other version.

To be clear, the Localisation and Thinning Principles are not statements of putative fundamental laws, but principles which state what is required from fundamental laws. What the specific laws will be is in part an empirical question. To make it more concrete, we can explore the theory in more detail by hypothetically assuming the truth of a specific scientific

proposal concerning the correlation between physical facts and the facts of conscious experience, namely that associated with the integrated information theory of consciousness, also known as 'IIT' (Oizumi, Albantakis, Tononi 2014).

IIT revolves around the notion of *integrated information* – the amount of which is represented by the Greek letter 'ϕ' – a concept which proponents of IIT attempt to give a mathematically precise definition of. We can avoid technical details here, except to say that, according to IIT, consciousness is correlated with *maximal ϕ*. That is to say, consciousness exists at the level at which there is most ϕ. If we want to know if some X is conscious, we need to ask two questions:

1. Does X have certain proper parts such that there is more ϕ in those proper parts than there is in X?
2. Is X a proper part of some greater whole such that there is more ϕ in that greater whole than there is in X?

If the answer to both (1) and (2) is 'no', then X is conscious, according to IIT. If there is more ϕ in the molecules making up my cup of coffee than there is in the cup of coffee considered as a whole, then IIT predicts that my cup of coffee is not conscious. If there is more ϕ in the cerebral cortex than there is either (i) in any of the neurons making it up, or (ii) in any whole of which the cerebral cortex is a part, then IIT predicts that the cerebral cortex is conscious. IIT not only tells us which physical entities are conscious, but also identifies the physical structures that correspond to the structures of human consciousness: roughly, they are those structures which support high levels of ϕ.

IIT is strictly speaking inconsistent with hybrid cosmopsychism, because on the latter view both the universe and I can be conscious at the same time. According to IIT, this is impossible: either the universe has more ϕ than me or vice versa; if the former, then I am not conscious as I am part of a greater whole which has more ϕ than I have; if the latter, then the universe is not conscious, as it has a part which has more ϕ than it itself has.[30]

However, we can consider a slightly modified form of IIT according to which *the existence and consciousness of local subjects* is correlated with maximal ϕ. Thus, we take it as given that the universe is conscious and hold that the principles of IIT tell us where there are conscious subjects over and above the universe subject and what kind of experience they have. This modified theory, call it IIT*, will be empirically indiscernible from original IIT, and thus in so far as we are judging matters on empirical grounds, this change ought not to make a difference.[31]

Assuming IIT* as the correct theory as to how the consciousness of local subjects is correlated with physical processes, our localisation and thinning laws will come out roughly as follows:

[30] I am assuming here that things that overlap spatially share parts. Perhaps this could be rejected, removing the inconsistency with standard IIT. I'm gratefully to an anonymous reviewer for pointing this out.

[31] IIT is at least partly supported by appeal to five 'axioms' of consciousness, justified on the basis of introspection. These are then translated into five corresponding postulates, which proponents of IIT allege that a system must exemplify in order to satisfy the axioms. It's the fifth postulate (the 'exclusion postulate,' according to which two conscious subjects cannot exist in a part-whole relationship) which is inconsistent with IIT*. The justification for the exclusion postulate seems to me decidedly weak (as, for example, argued by David Chalmers http://consc.net/slides/iit.pdf), so I'm not too concerned that IIT* violates it.

- Localisation Law: For any proper part of the universe P, P is a local subject iff (P has more ϕ than any proper parts of P) and (P has more ϕ than any whole of which P is a proper part).
- Thinning Law: For any conscious proper part of the universe P, those aspects of cosmic experience located within P that support high levels of ϕ are transferred to P.[32]

If IIT* is the correct theory of how the consciousness of local subjects is correlated with physical processes, then the above two principles will predict that humans and non-human

[32] Further complexities may be introduced when it comes to the *persistence* of emergent subjects. I am grateful to Bradford Saad for prompting me to think about this. Suppose we are working with an *endurantist* model of persistence, such that a subject is wholly present at each moment at which it exists. We could say that a subject continues to exist so long as its parts compose a system with maximal ϕ. But there are tricky cases. Consider, for example, the following possibility: at T1 certain atoms compose a system of maximal ϕ, at T2 all but one of those atoms cease to compose a system of maximal ϕ, whilst at the same moment one of those atoms joins some other atoms to compose a system which at that moment achieves maximal ϕ. Should we say that the subject that was wholly present at T1 is also wholly present at T2? Or has a new subject been brought into existence? Or suppose at T1 certain atoms compose a system of maximal ϕ, and at T2 50% of those atoms compose one system with maximal ϕ whilst 50% of those atoms compose a distinct system of maximal ϕ. Presumably at T2 we have a situation in which there are two distinct subjects, X and Y. Should we say that one of these subjects, either X or Y, is identical to the subject that existed at T1? If so, which one? X and Y cannot both be identical with the subject that existed at T1, as in that case (by the transitivity of identical) X and Y would be identical with each other, which *ex hypothesi* they are not. Or should we say that the subject that existed at T1 ceases to exist and two brand new subjects come into existence? There may be no principled way of decided which of these options is to be preferred. Matters are simpler if we adopt a four-dimensionalist view, according to which subjects are temporal parts of spacetime worms. We can then say that there are spacetime worms corresponding to all of the options outlined above, leaving us the conceptual choice of which temporal parts we want to group together and label as parts of a single person. The downside of this option, as I have explored in Goff 2017: Ch. 10, is that subjects of experience – the entities which have consciousness non-derivatively – do not persist through time. Why not identify subjects with spacetime worms? Perhaps there's something that it's like to be a spacetime worm, but it would be a very strange kind of consciousness, not at all like the kind of consciousness we pre-theoretically associate with human beings. At best, we could say that spacetime worms derivatively instantiate normal human consciousness (i.e. there is a spacetime worm that 'has human experience' in the sense of having temporal parts with human experience) and is thereby a human subject in a derivative sense. In any case, developing a detailed account of the persistence of subjects/persons will go hand in hand with developing a specific empirical account of the emergent laws (these issues may, for example, give us grounds for doubting IIT), whereas in this paper I am merely sketching the general view.

animals have the kind of conscious experience they actually do have. Again, this is exactly the result we want.[33]

IV

Some readers might have been wondering why the view under consideration needs to be construed as a form of *cosmo*psychism? Couldn't we conceive of a *micro*psychist version according to which in the beginning, as it were, there were only conscious particles, and then later macro-level conscious subjects emerge which take on some of the phenomenal properties that previously belonged to conscious particles?[34] The trouble with hybrid *micro*psychism – as opposed to hybrid *cosmo*psychism – is that it suffers from an especially pernicious version of what Daniel Stoljar dubbed 'the structural mismatch problem,'[35] the difficulty panpsychism has in accounting for the apparent mismatch between the structure of the brain and the structure of human consciousness.

[33] One might worry that the universe is going to end up with no experiences, given that there are maxima of ϕ covering the entire structure of the universe down to the microlevel (e.g. protons are maxima of ϕ). I am grateful to an anonymous reviewer for raising this worry. However, given that the fundamental fields borne by the universe cover every region of space, the universe will still be left with the consciousness underlying the regions of those fields in between protons and other maxima of ϕ.

[34] This view would be similar, but not the same as, the fusion view of Hedda Hassel Mørch (2014) and William Seager (2016), according to which particles either cease to exist in the process of fusing into an emergent subject (Seager 2016) or come to exist as entities derivative on the emergent whole (Mørch 2014). On the fusion view, *all* properties instantiated in the region occupied by the emergent whole would come to be borne by the emergent subject itself, whereas the hybrid view envisages that emergent and non-emergent subjects *share* properties in the region they occupy. Also, there is no requirement on the fusion view that properties are left *unchanged* by the emergence of a new subject to bear them, whereas this is an essential part of the hybrid view. If the proponent of the fusion view holds that the relevant properties are *not* changed by coming to be borne by an emergent subject, then they will also face the structural mismatch problem outlined below. If, on the other hand, they hold that the relevant properties *are* changed by coming to be borne by an emergent subject, then they will face the empirical challenge of explaining why this change in the properties of basic physics does not result in systems involving emergent subjects obeying different laws of nature to systems not involving emergent subjects (this second problem is similar to the empirical concerns for dualism/strong emergence discussed above).

[35] Nagasawa & Wager (2016) also hold that cosmopsychism can help avoid the structural mismatch problem.

If, as proposed in the hybrid micropsychist view outlined above, my consciousness is wholly formed from the *intrinsic* phenomenal properties of particles coming to be borne by me rather than the particles, it follows that the *relations* between those particles do not feature in my experience. But it is an obvious empirical fact that the structure of my consciousness corresponds to high-level structures in the brain, and that the latter structures in some sense reflect relationships between particles.[36] On the micropsychist view under consideration, relations between particles may play a role in determining *when* emergence happens, and perhaps *which* intrinsic phenomenal properties are taken on by the emergent subject, but the relations cannot themselves feature in the consciousness of the emergent subject. The only kind of consciousness that could be formed in this way is a structureless aggregate of a large number of 'pinpricks' of consciousness all experienced at once.

But couldn't the phenomenal properties of the particles retain their spatial relationships when they are inherited by my mind, thus ensuring that their relational structure features in my experience? I think this proposal conflates an *arrangement of experiences* with *an experience of arrangement*. Suppose I am experiencing six phenomenal properties – pain, itchiness, excitement, hunger, an image of red, a buzzing sound, and a sense of disappointment – and the six brain states instantiating those phenomenal properties happen to be arranged in a circle. It doesn't follow that I will have an experience of those phenomenal qualities arranged in a circle. The strong emergentist element of hybrid micropsychism ensures that micro-level experiences become unified in a strongly emergent

[36] I use the word 'reflect' so as to avoid committing to these structures being dependent on more fundamental facts about particles, which cosmopsychists deny.

subject, but the explanatory gap between microphysical structure and the structure of experience remains wide as ever.

Hybrid cosmopsychism, in contrast avoids the structural mismatch problem by starting not with a large number of pinpricks of consciousness but with a unified experience with rich structure, structure that corresponds not only to particles but also to the relationships between particles. The process of thinning out merely extracts a simpler structure from that more complex structure.[37]

V

A theory of consciousness is subject to two constraints:

> *The External Constraint* – The theory should be consistent with the empirical data (this constraint is in the background of causal exclusion worries).
>
> *The Internal Constraint* – The theory should avoid explanatory gaps, by ensuring that one could in principle deduce the facts of animal consciousness from the facts from which they strongly/weakly emerge (where we include in the emergence base, the fundamental laws, if there are any, which underwrite the emergence).

The hardness of the problem of consciousness can be traced to the difficulty of satisfying both of these constraints at the same time. In hybrid cosmopsychism, we find a theory that

[37] I don't mean to imply by this that particles are fundamental: see last footnote.

is up to the job. It satisfies the internal constraint via its appeal to basic laws determining the emergence of local subjects and specifying which phenomenal properties they take on from the universe (the strong emergentist component). It satisfies the external constraint because the emergence of local subjects does not introduce new fundamental properties (the weak emergentist component). No doubt there are further problems to address and further details to be worked out. However, in broad outline, hybrid cosmopsychism is a theory of consciousness we should take very seriously indeed.

References

Albahari, M. (2019). Beyond cosmopsychism and the great I am: How the world might be grounded in universal "advaitic" consciousness. In Seager (pp. 119-143).

Alter, T. & Nagasawa, Y. (2015). (Eds.) *Consciousness and the Physical World*, New York: Oxford University Press.

Baars, Bernard J. (2002). The conscious access hypothesis: Origins and recent evidence, *Trends in Cognitive Sciences*, *6*, 47-52. doi: 10.1016/s1364-6613(00)01819-2

Balog, K. (1999). Conceivability, possibility, and the mind-body problem, *Philosophical Review*, 108, 497-528. doi: 10.2307/2998286

Bird, A. (2007). *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press.

Brüntrup, G. & Ludwig J. (2016). *Panpsychism: Contemporary Perspectives*. New York: Oxford University Press.

Campbell, K. (1976). *Metaphysics: An Introduction.* Belmont: Dickenson.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Chalmers, David J. (2002). Consciousness and its place in nature. In D. J. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 247-272). New York: Oxford University Press.

Chalmers, D. J. (2006). Strong and weak emergence. In P. Davies & P. Clayton (Eds.) *The Re-Emergence of Emergence: The Emergentist Hypothesis From Science to Religion* (pp. 244-254). New York: Oxford University Press.

Chalmers, D. J. (2015). 'Panpsychism and panprotopsychism,' in Alter & Nagasawa 2015.

Chalmers, D. J. (2016) The combination problem for panpsychism,' in G. Brüntrup & L. Jaskolla 2016 (pp. 179-214).

Coleman, S. (2015). Neuro-cosmology. In P. Coates & S. Coleman (Eds.) *Phenomenal Qualities: Sense, Perception and Consciousness*, Oxford: Oxford University Press.

Coleman, S. (2016). Panpsychism and neutral monism: How to make up one's mind. In Brüntrup & Jaskolla (pp. 249-282).

Diaz-Leon, Esa (2008). Defending the phenomenal concept strategy, *Australasian Journal of Philosophy*, *86*, 597-610. doi: 10.1080/00048400802215638

Ellis, B. (2001). *Scientific Essentialism*. Cambridge: Cambridge University Press.

Ellis, B. (2002). *The Philosophy of Nature: A Guide to the New Essentialism*. Montreal: McGill-Queen's University Press.

Gibb, S. (2015). Defending dualism. *Proceedings of the Aristotelian Society, 115*, 131-146. doi: 10.1111/j.1467-9264.2015.00388.

Goff, Philip (2009) 'Why panpsychism doesn't help explain consciousness,' *Dialectica*, 63: 3, 289-311.

Goff, Philip (2015) 'Against constitutive Russellian monism,' in Alter, T. and Nagasawa, Y. (Eds.) *Russellian Monism*, Oxford University Press.

Goff, Philip (2016) 'Fundamentality and the mind-body problem,' *Erkenntnis* 8: 4, 881-898 (2016).

Goff, Philip (2016) 'Bertrand Russell and the Problem of Consciousness,' in J. Tartaglia & S.

Leach (Eds.) *Consciousness and the Great Philosophers,* Routledge.

Goff, Philip (2017a) *Consciousness and Fundamental Reality*, Oxford University Press.

Goff, Philip (2017b) 'Panpsychism,' in S. Schneider and M. Velmans (Eds.) *Blackwell Companion to Consciousness, 2nd Ed*, Blackwell.

Goff, Philip (2017c) 'Panpsychism is crazy, but it's also most probably true,' *Aeon*.

Goff, Philip (2017d) 'Do electrons dream of electric sheep?' *The Forum* (philosophy blog at LSE).

Goff, Philip, Seager, William, & Allen-Hermanson, Sean (2017/2022) 'Panpsychism' (radical update to existing entry), *Stanford Encylopedia of Philosophy* (2017). As single author, I radically rewrote the earlier version in 2017.

Goff, Philip (2018) 'Is the universe a conscious mind?' *Aeon*.

Goff, Philip (2019a) *Galileo's Error: Foundations of a New Science of Consciousness*, Pantheon.

Goff, Philip (2019b) 'Panpsychism' *Routledge Encyclopedia of Philosophy.*

Goff, Philip (2019c) 'Did the universe design itself?' *International Journal for the Philosophy*

of Religion 85: 99-122.

Goff, Philip (2019d) 'Grounding, analysis, and Russellian monism' S. Coleman (Ed.) *The Knowledge Argument Then and Now*, Cambridge University Press.

Goff, Philip (2020) 'Russellian monism,' (co-written with Sam Coleman) in U. Kriegal (Ed.) *Oxford Handbook of the Philosophy of Consciousness*.

Goff, Philip (2021) 'Cosmopsychism, micropsychism and the grounding relation,' in W. Seager (Ed.) *Routledge Handbook of Panpsychism*, Routledge.

Goff, Philip (forthcoming) *Why? The Purpose of Existence*, Oxford University Press.

Heil, J. (2003) *From an Ontological Point of View*, Oxford: Clarendon Press.

Lowe, E. J. (2006) *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*, Oxford: Oxford University Press.

Martin, C. B. (1980) Substance Substantiated. *Australasian Journal of Philosophy*, 58: 1, 3–10.

Kim, J. (1989). Mechanism, purpose, and explanatory exclusion. *Philosophical Perspectives, 3,* 77-108. doi: 10.2307/2214264

Loar, B. (1990/1997). Phenomenal states. Originally published in *Philosophical Perspectives 4*, 81-108. Reprinted in substantially revised form in N. Block, O. Flanagan, & G. Guüzeldere (Eds.) *The Nature of Consciousness: Philosophical Debates* (pp. 597-616). Cambridge, MA: MIT Press. doi: 10.2307/2214188

Lowe, E. J. (2006). *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Oxford: Oxford University Press.

Malcolm, N. (1968). The conceivability of mechanism. *Philosophical Review, 77,* 44-72. doi: 10.2307/2183182

Mathews, F. (2011). Panpsychism as Paradigm? In M. Blamauer (Ed.), *The Mental as Fundamental* (pp. 141-156), (Process Thought Series; edited by N. Rescher, J. Seibt & M. Weber). Heusenstamm: Ontos Verlag.

McClelland, T. (2013). The neo-Russellian ignorance hypothesis: A hybrid account of phenomenal consciousness. *Journal of Consciousness Studies 20,* 125-151.

McLaughlin, B. (1994). Epiphenomenalism. In S. Guttenplan (Ed.), *A Companion to the Philosophy of Mind* (pp. 277-288). Oxford: Blackwell.

Mendelovisi, A. (2018). *The Phenomenal Basis of Intentionality*. New York: Oxford University Press.

Molnar, G. (2003). *Powers: A Study in Metaphysics*. Oxford: Oxford University Press.

Mørch, H. H. (2014). *Panpsychism and Causation: A New Argument and a Solution to the Combination Problem*. Ph.D. thesis, University of Oslo.

Mørch, H. H. (2018). Is Consciousness Intrinsic? A Problem for the Integrated Information Theory. *Journal of Consciousness Studies, 26,* 133-162. Mørch, H. H. (2023) *Non-Physicalist Theories of Consciousness*, Cambridge: Cambridge University Press.

Nagasawa, Y. & Wager, K. (2016). Panpsychism and Priority Cosmopsychism. In Brüntrup & Jaskolla (pp. 113-129).

Nida-Rümelin, M. (2013). The Argument for Subject Body Dualism from Transtemporal Identity Defended, *Philosophy and Phenomenological Research, 86,* 702-714. doi: 10.1111/phpr.12019

O'Connor, T. (2018). For emergent individualism. In J. J. Loose, A. J. L. Menuge, J. P. Moreland (Eds.), *The Blackwell Companion to Substance Dualism* (pp. 247-272). Oxford: Wiley Blackwell. https://doi.org/10.1002/9781119468004.ch24

Oizumi, M., Albantakis, L., Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLOS Computational Biology 10*, 1-25. doi: 10.1371/journal.pcbi.1003588

Papineau, D. (2000) The Rise of Physicalism. In C. Gillett & B. M. Loewer (Eds.), *Physicalism and its Discontents*, Cambridge University Press.

Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Clarendon Press.

Robinson, H. (1982). *Matter and Sense.* Cambridge: Cambridge University Press.

Robinson, W. S. (2019) *Epiphenomenal Mind: An Integrated Outlook on Sensations, Beliefs, and Pleasure*, Routledge.

Roelofs, L. (2019). *Combining Minds: How to Think about Composite Subjectivity*. New York: Oxford University Press.

Russell, B. (1927). *The Analysis of Matter*. London: Kegan Paul.

Schaffer, J. (2010). Monism: The priority of the whole, *Philosophical Review, 119,* 31-76. doi: 10.1215/00318108-2009-025

Schaffer, J. (2017). The ground between the gaps, *Philosophers' Imprint, 17,* 1-26.

Seager, W. E. (1995). Consciousness, information, and panpsychism, *Journal of Consciousness Studies, 2,* 272-88.

Seager, W. E. (2016). Panpsychism Infusion. In Brüntrup & Jaskolla (pp. 229–248).

Seager, W. E. (2019). *The Routledge Handbook of Panpsychism*, New York, Oxford: Routledge.

Shani, I. (2015). Cosmopsychism: A Holistic Approach to the Metaphysics of Experience, *Philosophical Papers, 44,* 389–417. doi: 10.1080/05568641.2015.1106709

Shani, I. & Keppler, J. (2018) Beyond Combination: How Cosmic Consciousness Grounds Ordinary Experience, *Journal of the American Philosophical Association 4,* 390-410. doi: 10.1080/05568641.2015.1106709.

Strawson, G. (2006). Realistic materialism: Why physicalism entails panpsychism, *Journal of Consciousness Studies, 13,* 10-11.