

# Panpsychism and Panprotopsychism<sup>1</sup>

David J. Chalmers

## 1 Introduction

Panpsychism, taken literally, is the doctrine that everything has a mind. In practice, people who call themselves panpsychists are not committed to as strong a doctrine. They are not committed to the thesis that the number two has a mind, or that the Eiffel tower has a mind, or that the city of Canberra has a mind, even if they believe in the existence of numbers, towers, and cities.

Instead, we can understand panpsychism as the thesis that some fundamental physical entities have mental states. For example, if quarks or photons have mental states, that suffices for panpsychism to be true, even if rocks and numbers do not have mental states. Perhaps it would not suffice for just one photon to have mental states. The line here is blurry, but we can read the definition as requiring that all members of some fundamental physical types (all photons, for example) have mental states.

For present purposes, the relevant sorts of mental states are conscious experiences. I will understand panpsychism as the thesis that some fundamental physical entities are conscious: that is, that there is something it is like to be a quark or a photon or a member of some other fundamental physical type. This thesis is sometimes called *panexperientialism*, to distinguish it from other varieties of panpsychism (varieties on which the relevant entities are required to think or reason, for example), but I will simply call it panpsychism here.

Panpsychism is sometimes dismissed as a crazy view, but this reaction on its own is not a serious objection. While the view is counterintuitive to some, there is good reason to think that any view of consciousness must embrace some counterintuitive conclusions. Furthermore, intuitions about panpsychism seem to vary heavily with culture and with historical period. The

---

<sup>1</sup> Forthcoming as the 2013 Amherst Lecture in Philosophy, and in a book on Russellian monism. I first presented this material at the Munich conference on panpsychism and emergence in June 2011. I am grateful to the audience there and also to audiences at Amherst, Bogazici, Charleston, Fordham, Notre Dame, Santiago, Stanford, and Wesleyan. I owe a special debt to Daniel Stoljar whose related work in “Two Conceptions of the Physical” greatly influenced this article. Thanks also to Torin Alter, Sam Coleman, Brian Garrett, Philip Goff, John Gregg, Bill Meacham, Daniel Stoljar, and Keith Turausky for comments on this article.

view has a long history in both Eastern and Western philosophy, and many of the greatest philosophers have taken it seriously. It is true that we do not have much direct evidence for panpsychism, but we also do not have much direct evidence against it, given the difficulties of detecting the presence or absence of consciousness in other systems. And there are indirect reasons, of a broadly theoretical character, for taking the view seriously.

In this article I will present an argument for panpsychism. Like most philosophical arguments, this argument is not entirely conclusive, but I think it gives reason to take the view seriously. Speaking for myself, I am by no means confident that panpsychism is true, but I am also not confident that it is not true. This article presents what I take to be perhaps the best reason for believing panpsychism. A companion article, “The Combination Problem for Panpsychism”, presents what I take to be the best reason for disbelieving panpsychism.

I call my argument the Hegelian argument for panpsychism. This is not because Hegel was a panpsychist. He seems to have been far from it, perhaps except insofar as he believed in a “world-soul” (which suggests a sort of cosmopsychism, the view that the world as a whole is conscious). Rather, my argument takes the dialectical form often attributed to Hegel: the form of thesis, antithesis, synthesis.<sup>2</sup>

In my Hegelian argument, the thesis is materialism, the antithesis is dualism, and the synthesis is panpsychism. The argument for the thesis is the causal argument for materialism (and against dualism). The argument for the antithesis is the conceivability argument for dualism (and against materialism). Synthesized, these yield the Hegelian argument for panpsychism. In effect, the argument presents the two most powerful arguments for and against materialism and dualism, and motivates a certain sort of panpsychism as a view that captures the virtues of both views and the vices of neither.

It turns out that the Hegelian argument does not support only panpsychism. It also supports a certain sort of *panprotopsychism*: roughly, the view that fundamental entities are *proto-conscious*, that is, that they have certain special properties that are precursors to consciousness and that can collectively constitute consciousness in larger systems. Later in the article, I will examine the relative merits of panpsychism and panprotopsychism, and examine problems that arise for both.

---

<sup>2</sup> I gather that in fact this dialectical form comes from Fichte, and that Hegel dismissed it as simplistic. Still, I will stay with the popular attribution.

## 2 Thesis and Antithesis: Materialism and Dualism

Our thesis is materialism (or physicalism): roughly, the thesis that everything is fundamentally physical. Our antithesis is dualism: roughly, the thesis that not everything is fundamentally physical, and the things that are not fundamentally physical are fundamentally mental. Our synthesis is panpsychism: very roughly, the thesis that everything is (or at least that some things are) fundamentally physical *and* fundamentally mental.

More specifically, we will be concerned with materialism and dualism about consciousness. Materialism about consciousness is the thesis that consciousness is fundamentally physical: that is, that truths about consciousness are grounded in the fundamental truths of a completed physics. Dualism about consciousness is the thesis that consciousness is not fundamentally physical: that is, that truths about consciousness are not grounded in the fundamental truths of a completed physics.

Grounding is a relation of metaphysical constitution. Truths about consciousness are grounded in physical truths if all truths in the first set obtain wholly in virtue of truths in the second set obtaining.<sup>3</sup> The intuitive idea behind materialism is that physical truths somehow add up to and yield truths about consciousness. This requires at least that there is a metaphysically necessary connection between these truths, in that it is impossible for a world to be physically like ours without that world being phenomenally like ours. Intuitively, once God created the entities of physics, consciousness came along for free.

We will be especially concerned with microphysical properties and with phenomenal properties. Microphysical properties are the fundamental physical properties characterized by a completed physics. Microphysical entities are the fundamental physical entities characterized by that physics. (Despite the name, it is not definitionally required that these entities be small.) Microphysical truths are positive truths about the instantiation of microphysical properties by microphysical entities. Here a positive truth is intuitively a truth about the properties that an entity has, rather than those that it lacks (for more on this, see Chalmers 2012). Macrophysical

---

<sup>3</sup> For more on the notion of grounding, see Schaffer 2009 and Fine 2012. The notion of grounding at play here is what is sometimes called “full grounding”, involving a “wholly in virtue of” relation, as opposed to “partial grounding”, which involves a “partly in virtue of” relation. The latter is inappropriate for defining materialism, as the definition would then allow a nonmaterialist view on which truths about consciousness obtain in virtue of physical truths along with some other nonphysical truths).

properties (entities, truths) are those that are grounded in microphysical properties (entities, truths). For ease of discussion, I will use the word “physical” to mean “microphysical” throughout what follows, sometimes using “microphysical” for explicitness.

Phenomenal (or experiential) properties are properties characterizing what it is to be a conscious subject. The most familiar phenomenal property is simply the property of phenomenal consciousness: an entity has this property when there is something it is like to be that entity. There are also many specific phenomenal properties, characterizing more specific conscious experiences. For example, phenomenal redness characterizes the distinct sort of conscious experience we have when we experience redness. An entity has the property of phenomenal redness when it has that sort of conscious experience. Phenomenal truths are positive truths about the distribution of phenomenal properties (that is, truths about what it is like to be various entities).

We can then say that materialism about consciousness is the thesis that all phenomenal truths are grounded in microphysical truths. Dualism about consciousness is the thesis that phenomenal truths are not all grounded in microphysical truths. In what follows, by “materialism” and “dualism” I mean materialism and dualism about consciousness.

We can put the conceivability argument against materialism (and for dualism) as follows. Here  $P$  is the conjunction of all microphysical truths about the universe, and  $Q$  is an arbitrary phenomenal truth (such as ‘I am conscious’).

- (1)  $P \& \sim Q$  is conceivable.
- (2) If  $P \& \sim Q$  is conceivable,  $P \& \sim Q$  is metaphysically possible.
- (3) If  $P \& \sim Q$  is metaphysically possible, materialism is false.

---

(4) Materialism is false.

Here we can say that a claim is conceivable when it is not ruled out a priori. So it is conceivable that there are mile-high unicycles, for example. A claim is metaphysically possible when it could have obtained: intuitively, when God could have created the world such that the claim would have been true. So it is plausibly metaphysically possible that there are mile-high unicycles.

Premise (1) here is supported by the conceivability of zombies: creatures microphysically identical to us without consciousness. Most people think that zombies do not actually exist, but there seems to be no a priori contradiction in the idea. Premise (2) is supported by general

reasoning about the relationship between conceivability and possibility. The thesis needs to be refined to accommodate various counterexamples due to Kripke and others, but I will stay with the simple thesis here.<sup>4</sup> Premise (3) is supported by the idea that if  $P \& \sim Q$  is metaphysically possible, then  $P$  does not metaphysically necessitate  $Q$ , so  $Q$  is not grounded in  $P$ , since grounding plausibly requires metaphysical necessitation. Here the intuitive idea is that if God could have created a world microphysically identical to our world but without consciousness, then the presence of consciousness involves new fundamental properties over and above those of physics, so materialism is false.

The conceivability argument is an epistemic argument against materialism, starting with an epistemological premise and proceeding to a metaphysical conclusion. There are other, closely related epistemic arguments. These include the knowledge argument, which starts from the premise that  $Q$  is not deducible from  $P$  and concludes that it is not grounded in  $P$ ; the explanatory argument, which starts from the premise that there is an explanatory gap between  $P$  and  $Q$  and concludes that there is an ontological gap; and the structure-dynamics argument, which starts from the premise that  $P$  can be analyzed in terms of structure and dynamics while  $Q$  cannot and concludes that  $Q$  is not grounded in  $P$ . Much of what I say will apply to all these arguments, but I will focus on the conceivability argument here.

Materialists do not just curl up and die when confronted with the conceivability argument and its cousins. Type-A materialists reject the epistemic premise, holding for example that zombies are not conceivable. Type-B materialists reject the step from an epistemic premise to an ontological conclusion, holding for example that conceivability does not entail possibility. Still, there are significant costs to both of these views. Type-A materialism seems to require something akin to an analytic functionalist view of consciousness, which most philosophers find too deflationary to be plausible. Type-B materialism seems to require a sort of brute necessity of a sort that is not found elsewhere and that is hard to justify. Of course some philosophers find these costs worth paying, or deny that these are costs. Still, I think that the argument makes at least a *prima facie* case against materialism.

That said, many materialists think that the conceivability argument against materialism (and for dualism) is countered by the causal argument against dualism (and for materialism). This argument runs as follows:

---

<sup>4</sup> For a much-elaborated version of the argument using two-dimensional semantics, see Chalmers 2009.

(1) Phenomenal properties are causally relevant to physical events.

(2) Every caused physical event has a full causal explanation in physical terms.

(3) If every caused physical event has a full causal explanation in physical terms, every property causally relevant to the physical is itself grounded in physical properties.

(4) If phenomenal properties are grounded in physical properties, materialism is true.

---

(5) Materialism is true.

Here we can say that a property is causally relevant to an event when instantiations of that property are invoked in a correct causal explanation of that event. For example, the high temperatures in Victoria were causally relevant to the Victorian bushfires. A full causal explanation of an event is one that characterizes *sufficient* causes of the event: causes that guarantee that the event will occur, at least given background laws of nature.

Premise (1) is supported by intuitive observation. My being in pain seems to cause my arm to move. If things are as they seem here, then the pain will also be causally relevant to the motion of various particles in my body. Premise (2) follows from a widely held view about the character of physics: physics is causally closed, in that there are no gaps in physical explanations of physical events. Premise (3) is a rejection of a certain sort of overdetermination. Given a full microphysical causal explanation of physical events, other causal explanations are possibly only when the factors involved in the latter are grounded in the factors involved in the former (as when we explain the motion of a billiard ball both in terms of another ball and in terms of the particles that make it up).<sup>5</sup> Any putative causal explanation that was not grounded in this way would involve causal overdetermination by independent events. Systematic overdetermination of this sort is widely rejected. Premise (4) is true by definition.

---

<sup>5</sup> Principles such as (3) are sometimes put with “is a physical property” instead of “is grounded in a physical property”. This amounts to an overly strong causal exclusion claim on which high-level events and their low-level grounds cannot both be causally relevant. Reflection on standard cases (Bennett 2003; Yablo 1992) suggests that constitutively connected events need not exclude each other as causes: these are cases of “benign overdetermination” as opposed to cases of “bad overdetermination”. Premise (3) excludes only cases of the latter sort.

Dualists do not just curl up and die when presented with the causal argument. Epiphenomenalists reject premise (1), holding that the claim that consciousness causes behavior is just an intuition and can be rejected. Interactionists reject premise (2), holding that physics leaves room for (and perhaps is positively encouraging to) causal gaps that consciousness might fill. Still, there are costs to both of these views. Epiphenomenalism is at least inelegant and requires special coincidences between conscious experiences and macrophysical events (utterances about consciousness, for example) that seem to reflect them. Interactionism requires a view of physics that would be widely rejected by most physicists, and that involves a large bet on the future of physics. Again, some dualists (including me in some moods) deny that these are costs or hold that the costs are worth paying. Still, I think there is at least a *prima facie* case against dualism here.

So we have a standoff. On the face of it, the conceivability argument refutes materialism and establishes dualism, and the causal argument refutes dualism and establishes materialism. It is time for a Hegelian synthesis.

### **3 Synthesis: Panpsychism**

Panpsychism, once again, is the thesis that some microphysical entities are conscious. For our purposes, it is useful to distinguish various more fine-grained varieties of panpsychism. To do this, we can first introduce some terminology.

Let us say that *macroexperience* is the sort of conscious experience had by human beings and other macroscopic entities (that is, entities that are not fundamental physical entities). Macroexperience involves the instantiation of *macrophenomenal properties*: properties characterizing what it is like to be humans and other macroscopic entities. Let us say that *microexperience* is the sort of conscious experience had by microphysical entities. Microexperience involves the instantiation of *microphenomenal properties*: properties characterizing what it is like to be microphysical entities.

If panpsychism is correct, there is microexperience and there are microphenomenal properties. We are not in a position to say much about what microexperience is like. I think we can be confident that it is very different from human experience, however. It is almost certainly much simpler than human experience. In the way that an experience of redness is much simpler than a stream of conscious thought, we can expect a quark's experience to be much simpler than

an experience of redness. To get far beyond generalities like this concerning microexperience, we would need a proper panpsychist theory of consciousness, which we are currently lacking.

*Constitutive panpsychism* is the thesis that macroexperience is (wholly or partially) grounded in microexperience. More or less equivalently, it is the thesis that macroexperience is constituted by microexperience, or realized by microexperience. On this view, macrophenomenal truths obtain in virtue of microphenomenal truths, in roughly the same sense in which materialists hold that macrophenomenal truths obtain in virtue of microphysical truths. To put things intuitively, constitutive panpsychism holds that microexperiences somehow add up to yield macroexperience. The view can allow that macroexperience is not wholly grounded in microexperience: for example, it might be grounded in microexperience along with certain further structural or functional properties.

Panpsychists need not be constitutive panpsychists. There is also nonconstitutive panpsychism, which holds that there is microexperience and macroexperience, but the microexperience does not ground the macroexperience. Nonconstitutive panpsychists will typically be *emergent panpsychists*, holding that macroexperience is strongly emergent from microexperience and/or from microphysics. One sort of emergent panpsychist might hold that there are contingent laws of nature that specify when certain microexperiences give rise to certain macroexperiences. Another might hold that there are laws of nature connecting microphysical properties to microphenomenal properties and macrophysical properties to macrophenomenal properties, without there being any constitutive connection between microphenomenal and macrophenomenal. Still, as we will see, nonconstitutive panpsychism inherits many of the problems of dualism, while it is constitutive panpsychism that offers hope for a Hegelian synthesis. So it is this view that I will focus on.

Like materialism, constitutive panpsychism comes in type-A and type-B varieties. Type-A constitutive panpsychism holds that there is an a priori entailment from microphenomenal truths to macrophenomenal truths, while type-B constitutive panpsychism holds that there is an a posteriori necessary entailment from microphenomenal truths to macrophenomenal truths. The type-B view inherits many of the problems of type-B materialism, so it is the type-A view that offers special hope for a Hegelian synthesis. So while I will talk of constitutive panpsychism, it will usually be the type-A version that I have especially in mind.

Another important variety of panpsychism is *Russellian panpsychism*. This view takes its name from Russell's insight, in *The Analysis of Matter* and other works, that physics reveals the



relational structure of matter but not its intrinsic nature. According to this view, classical physics tells us a lot about what mass does—it resists acceleration, attracts other masses, and so on—but it tells us nothing about what mass intrinsically is. We might say that physics tells us what the mass role is, but it does not tell us what property plays this role.

Here we can say that *quiddities* are the fundamental categorical properties that play the fundamental roles specified in physics. Alternatively, we can say that quiddities are the categorical bases of the microphysical dispositions characterized in physics. We can stipulate in addition that quiddities are distinct from the roles or the dispositions themselves. A view on which there are only role or dispositional properties, and no distinct properties playing those roles or serving as the basis for the dispositions, is a view on which there are no quiddities.

It is not obvious that there must be quiddities. There are respectable structuralist or dispositionalist views of physics on which physics involves just structure or dispositions all the way down. Still, many find these views objectionable, because they seem to yield a world devoid of substance or qualities—Russell said that on views like these “all the things in the world will merely be each others’ washing” (*The Analysis of Matter*, p. 325). And whether or not one accepts these objections, it is certainly not obvious that there are no quiddities. On the face of it, a worldview that postulates quiddities is perfectly coherent, and there is little clear evidence against it.

Russellian panpsychism is the view that some quiddities are microphenomenal properties. This view requires that there are quiddities—distinct properties that play the mass role, the charge role, and so on—and that at least some of these quiddities are phenomenal. For example, perhaps the property that plays the mass role is a certain phenomenal property. (Or better, as mass is really a quantity: the quantity that plays the mass role is a certain phenomenal quantity.) The Russellian panpsychist addresses two metaphysical problems—what is the place of phenomenal properties in nature, and what are the intrinsic properties underlying physical structure?—and in effect answers both of them at once. Fundamental phenomenal properties play fundamental microphysical roles and underlie fundamental microphysical structure.

Panpsychists need not be Russellian panpsychists. There is also non-Russellian panpsychism, according to which there are microphenomenal properties that do not play microphysical roles. Perhaps there are numerous microphenomenal properties quite distinct from the properties involved in the microphysical network, for example. Still, non-Russellian

panpsychism faces obvious problems with mental causation, while Russellian panpsychism that offers hope for a Hegelian synthesis. So it is this view that I will focus on.

In particular, I will focus on constitutive Russellian panpsychism. On this view, microphenomenal properties serve as quiddities, playing the roles associated with microphysical properties, and also serve as the grounds for macrophenomenal properties. That is, microexperience constitutes macroexperience while also playing microphysical roles. On this view, one could think of the world as fundamentally consisting in fundamental entities bearing fundamental microphenomenal properties, where these microphenomenal properties are connected to each other (and perhaps to other quiddities) by fundamental laws with the structure that the laws of physics describe. All this microphenomenal structure also serves to constitute the macrophenomenal realm, just as microphysical structure serves to constitute the macrophysical realm.

I think that constitutive Russellian panpsychism is perhaps the most important form of panpsychism, precisely because it is this form that promises to avoid the problems of physicalism and dualism and to serve as a Hegelian synthesis. In particular, one can argue that this view avoids both the conceivability argument against physicalism and the causal argument against dualism.

To assess this matter, we first need to assess a delicate question: is constitutive Russellian panpsychism a form of materialism, a form of dualism, or neither? This question turns on the answer to another delicate question: are quiddities physical properties? If quiddities are physical properties, then constitutive Russellian panpsychism entails that microphenomenal properties are physical properties, and that macrophenomenal properties are constituted by physical properties, so that materialism is true. If quiddities are not physical properties, however, then macrophenomenal properties will be constituted by nonphysical properties, and a form of dualism will be true.

To answer this question, it is useful to make a distinction. We can say that *narrowly physical* properties are microphysical role properties, such as the dispositional property associated with having a certain mass, or the second-order property of having a property that plays the mass role.<sup>6</sup> We can say that *broadly physical* properties are physical role properties along with any

---

<sup>6</sup> The distinction between narrowly and broadly physical properties is closely related to Stoljar's distinction between t-physical properties (properties invoked by physical theory) and o-physical properties (intrinsic properties of physical objects), but it is not the same

properties that realize the relevant roles: categorical bases for the mass dispositions, first-order properties that play the mass role.

In effect, narrowly physical properties include structural properties of microphysical entities but exclude quiddities, while broadly physical properties include both structural properties and quiddities. Here a structural property is one that can be fully characterized using structural concepts alone, which I take to include logical, mathematical, and nomic concepts, perhaps along with spatiotemporal concepts (see Chalmers 2003 and 2012 for much more discussion). If one uses a Ramsey sentence to characterize fundamental physics, it is plausible that one can do so using structural concepts alone. At the same time, if there are quiddities, it is plausible that they (like phenomenal properties) cannot be fully characterized in structural terms.

We can then say that quiddities are not narrowly physical, but they are broadly physical. There is more to say here, particularly concerning just how we should construe the relation between quiddities and ordinary physical properties such as mass, but I will leave this for the next section.

With this distinction made, the question of whether quiddities are physical properties becomes something of a verbal question. One can use the term “physical” to cover only narrowly physical properties or to cover broadly physical properties, and the choice between these usages is a verbal matter. Some may think that there is a stronger case for one usage or the other, but little of substance turns on this.

The same applies to the question of whether constitutive Russellian panpsychism is physicalism. We can distinguish narrow physicalism, which holds that phenomenal truths are grounded in narrowly physical truths, from broad physicalism, which holds that phenomenal truths are grounded in broadly physical truths. Narrow physicalism entails broad physicalism, but broad physicalism may not entail narrow physicalism. In particular, constitutive Russellian panpsychism is incompatible with narrow physicalism, but it is a form of broad physicalism.

---

distinction. For a start, given a view on which ‘mass’ refers to a quiddity that plays the mass role, then mass will be t-physical (assuming a property is invoked by physical theory iff it is referred to by an expression of that theory) but it will not be narrowly physical. And given a view on which physical objects have epiphenomenal intrinsic properties that are not those invoked by physical theories and that are not quiddities, these properties will be o-physical but will not be broadly physical. For related reasons (discussed later), I think the broad/narrow distinction is better-suited than the t-/o- distinction to do the work that Stoljar wants the latter to do. Note that in Chalmers 2010 (p. 192) I use the broad/narrow terminology to mark a different distinction.

Once again, any dispute over whether narrow or broad physicalism is really physicalism will be something of a verbal dispute. Instead, constitutive Russellian panpsychism falls into a penumbral area that might be counted either way. This is a promising area for a Hegelian synthesis.

How does constitutive Russellian panpsychism fare with respect to the conceivability argument against physicalism? Once we have the distinction between narrowly and broadly physical truths in place, we can distinguish two different versions of the argument. One version construes  $P$  as the conjunction of all positive narrowly physical truths, takes as a premise that the corresponding version of  $P \& \sim Q$  is conceivable, and concludes that narrow physicalism is false. The other does the same for broadly physical truths and broad physicalism.

To assess these arguments, we can distinguish two different sorts of zombies: narrowly physical duplicates of us without consciousness, and broadly physical duplicates of us without consciousness. We can call the first group *structural zombies*, since they duplicate just our relational physical structure. We can call the second group *categorical zombies*, since they also duplicate the underlying categorical properties.<sup>7</sup>

It is plausible that when we typically conceive of zombies, we are really conceiving of structural zombies. We hold physical structure fixed, but we do not make any effort to hold quiddities fixed, since we have no idea what the quiddities are. This standard zombie intuition provides good reason to think that structural zombies are conceivable, but little reason to think that categorical zombies are conceivable. If this is right, adding the conceivability-possibility premise at best establishes the possibility of structural zombies but not of categorical zombies. This is a happy result for (type-A) constitutive Russellian panpsychists, who hold that categorical zombies are not conceivable and not possible.

The upshot of this is that the standard considerations about conceivability can be used at most to undermine narrow physicalism, but not broad physicalism. So these considerations have no force against constitutive Russellian panpsychism, which is a version of the latter but not the former. So this view evades at least one horn of the Hegelian dilemma.

What about the other horn: the causal argument against dualism? Here it is useful to first reflect on the causal role of experience under constitutive Russellian panpsychism. According to

---

<sup>7</sup> Structural and categorical zombies are closely related to the t-zombies (t-physical duplicates without consciousness) and o-zombies (o-physical duplicates without consciousness) discussed by Stoljar (2001a). As before I think the broad/narrow distinction is more crucial than the t-/o- distinction here.

Russellian panpsychism, microphenomenal properties certainly play a causal role in physics. They are the properties that play the most fundamental causal roles in physics: the mass role, the charge role, and so on. A microphenomenal property that plays the mass role is causally responsible for attracting other entities, and so on. This causation does not involve any violation of the laws of physics. Instead, this sort of causation underlies the laws of physics.

At the same time, constitutive panpsychism allows that macroexperience can inherit causal relevance from microexperience. This is an instance of the general claim that constituted properties can inherit causal relevance from constituting properties. For example, a billiard ball can inherit causal relevance from that of the particles that make it up. I think this is the lesson of much recent discussion of causal exclusion between the microscopic and macroscopic levels: when entities at this level are constitutively connected, there need be no causal exclusion. The moral that applies to the microphysical and the macrophysical also applies to the microphenomenal and the macrophenomenal, if they are constitutively connected.

It follows that constitutive Russellian panpsychism is compatible with a robust causal role for both microexperience and macroexperience. Given that microexperience is causally relevant (as Russellian panpsychism suggests), and that microexperience constitutes macroexperience (as constitutive panpsychism suggests), we can expect that macroexperience will be causally relevant too.

What of the causal argument? Here again we need to distinguish versions of the argument. One version of the argument invokes the causal closure of the broadly physical to argue that phenomenal properties are grounded in broadly physical properties. The premises of this version of the argument are all plausible, and the constitutive Russellian panpsychist can happily accept its conclusion. Another version invokes the causal closure of the narrowly physical to argue that phenomenal properties are grounded in narrowly physical properties. Here the constitutive Russellian panpsychist must reject the conclusion, but fortunately they can easily reject premise 2. A full causal explanation of narrowly physical events will involve broadly physical properties; a causal explanation wholly in terms of narrowly physical properties is incomplete. This is to say that on a view where there are quiddities, the broadly physical domain may be causally closed, but the narrowly physical domain will not be.<sup>8</sup>

---

<sup>8</sup> Alternatively, the constitutive Russellian panpsychist can accept premise (2) asserting the causal closure of the narrowly physical, while rejecting premise (3). They can hold that the narrowly physical explanation is itself grounded in a broadly physical explanation, so that

The upshot is that the causal argument can be used at best to establish broad physicalism and not narrow physicalism. This is once again a happy result for the constitutive Russellian panpsychism, as it is a version of the former but not the latter. So this view evades the second horn of our Hegelian dilemma.

We can combine our analysis of the two arguments as follows. The conceivability argument refutes narrow physicalism but is compatible with broad physicalism. The causal argument establishes broad physicalism but does not establish narrow physicalism. The arguments put together yield the Hegelian argument for the conjunction of broad physicalism with the denial of narrow physicalism. This is the ground occupied by constitutive Russellian panpsychism.<sup>9</sup>

It is worth noting that nonconstitutive and non-Russellian panpsychism do not evade the Hegelian dilemma. Both of these views are incompatible with broad physicalism, and so are vulnerable to the causal argument for broad physicalism. On nonconstitutive panpsychism, even if microexperience is causally relevant, macroexperience will lie outside the broad physical network, so it will lead to epiphenomenalism, interactionism, or overdetermination. On nonRussellian panpsychism, it is hard to see how even microphenomenal properties can be causally relevant, and the same trilemma ensues. Among versions of panpsychism, only constitutive Russellian panpsychism promises to serve as a Hegelian synthesis.

## 4 Antithesis: Panprotopsychism

It is a familiar point in the pseudo-Hegelian dialectic that every synthesis is confronted by a new antithesis and followed by a new synthesis. Our Hegelian synthesis above is panpsychism. But it turns out that another view can also escape the original Hegelian dilemma: constitutive Russellian panprotopsychism.

---

these explanations are not independent and a bar on overdetermination does not render them incompatible. The case for premise (3) tacitly assumes that physical explanations do not themselves have further grounds; but on a Russellian view, narrowly physical explanations have further grounds.

<sup>9</sup> The Hegelian argument could in principle be formalized as a six-premise argument that uses the three premises of the conceivability argument (with “physical” disambiguated to mean narrowly physical) and the three premises of the causal argument (with “physical” to mean broadly physical) to establish the conjunction of broad physicalism with the denial of narrow physicalism. An argument structure along these lines is at play in Stoljar (2001b, section 4), with the main differences being that Stoljar invokes the knowledge argument rather than the conceivability argument, uses the o-/t- distinction where I use the broad/narrow distinction, and rejects panpsychism.

Recall that panprotopsychoism is the view that fundamental physical entities are proto-conscious. In more detail, let us say that *protophenomenal* properties are special properties that are not phenomenal (there is nothing it is like to have a single protophenomenal property) but that can collectively constitute phenomenal properties, perhaps when arranged in the right structure. Panprotopsychoism is then the view that some fundamental physical entities have protophenomenal properties.

One might worry that any non-panpsychist materialism will be a form of panprotopsychoism. After all, non-panpsychist materialism entails that microphysical properties are not phenomenal properties and that they collectively constitute phenomenal properties. This is an undesirable result. The thought behind panprotopsychoism is that protophenomenal properties are special properties with an especially close connection to phenomenal properties. To handle this, one can unpack the appeal to specialness in the definition by requiring that (i) protophenomenal properties are distinct from structural properties and (ii) that there is an a priori entailment from truths about protophenomenal properties (perhaps along with structural properties) to truths about the phenomenal properties that they constitute. This excludes ordinary type-A materialism (which grounds phenomenal properties in structural properties) and type-B materialism (which invokes an a posteriori necessary connection). From now on I will understand protophenomenal properties this way, and will understand panprotopsychoism accordingly.<sup>10</sup>

I have occasionally heard it said that panprotopsychoism can be dismissed out of hand for the same reason as materialism. According to this objection, the epistemic arguments against materialism all turn on there being a fundamental epistemic (and therefore ontological) gap between the nonphenomenal and the phenomenal: there is no a priori entailment from nonphenomenal truths to phenomenal truths. If this were right, the gap would also refute panprotopsychoism. I do not think that this is right, however. The epistemic arguments all turn on a more specific gap between the physical and the phenomenal, ultimately arising from a gap between the structural (or the structural/dynamical) and the phenomenal. We have principled reasons to think that phenomenal truths cannot be wholly grounded in structural truths. But we

---

<sup>10</sup> What about type-B views that appeal to quiddities that satisfy (i) but not (ii)? Some such views may nevertheless have a “panprotopsychoist” flavor, perhaps because of the special flavor of the quiddities they appeal to, while others (say, the view advocated by Papineau 2002) seem to lack that flavor. A line between these views is hard to draw, so for present purposes I count none of them as panprotopsychoism.

have no correspondingly good reason to think that phenomenal truths cannot be wholly grounded in nonphenomenal (and nonstructural) truths, as panprotopsychism suggests.

It is true that we do not have much idea of what protophenomenal properties are like. For now they are characterized schematically, in terms of their relation to phenomenal properties. A fuller account will have to wait for a full panprotopsychist theory, though I will speculate about one sort of protophenomenal property toward the end of this article. But our ignorance about protophenomenal properties should not be mistaken for an objection to the truth of panprotopsychism.

*Constitutive* panprotopsychism is roughly the thesis that macroexperience is grounded in the protophenomenal properties of microphysical entities. That is, all phenomenal truths are grounded in protophenomenal truths concerning these entities. As before, constitutive panprotopsychism could in principle come in type-A and type-B varieties, but the definition of specialness above in effect restricts it to the type-A version (a priori entailment from protophenomenal truths to macrophenomenal truths), which is in any case the relevant version for our purposes. *Russellian* panprotopsychism is the thesis that some quiddities are protophenomenal properties. For example, perhaps protophenomenal properties play the mass role or the charge role.

Nonconstitutive and nonRussellian panprotopsychism are coherent theses as protophenomenal properties are defined above (at least if we set aside the specialness clause): perhaps protophenomenal properties only constitute some macroexperiences, and perhaps they do not serve as quiddities. As with panpsychism, however, the Hegelian motivations for panprotopsychism strongly favor (type-A) constitutive Russellian panprotopsychism, so it is this view on which I will concentrate.

Constitutive Russellian panprotopsychism, like constitutive Russellian panpsychism, is a form of broad physicalism without narrow physicalism. It therefore escapes the Hegelian dilemma in just the same way. Constitutive Russellian panpsychists will reply to the conceivability argument by saying that structural zombies are conceivable but categorical zombies are not. They will reply to the causal argument by saying that fundamental protophenomenal properties are causally relevant in virtue of playing microphysical roles, and that macrophenomenal properties inherit causal relevance from protophenomenal properties in virtue of being grounded in them. In this way it slips through the horns of the Hegelian dilemma.



## 5 Synthesis: Russellian monism

Given panpsychism as thesis and panprotopsyism as antithesis, there is a natural synthesis that subsumes them both. This synthesis is Russellian monism. We can understand Russellian monism as the conjunction of broad physicalism with the denial of narrow physicalism. On this view, structural properties in physics do not constitute consciousness, but quiddities (perhaps along with structure) constitute consciousness. The view is Russellian because of the appeal to quiddities and their connection to mentality. It is a sort of monism because the world on this view consists in quiddities connected by laws of nature.

It is easy to see that both constitutive Russellian panpsychism and constitutive Russellian panprotopsyism are forms of Russellian monism. In fact, Russellian monism is equivalent to the disjunction of the two. According to Russellian monism, all conscious experience is grounded in structure plus quiddities, but not in structure alone. Given the definition of protophenomenal properties above, this thesis is equivalent to the thesis that some quiddities are either phenomenal or protophenomenal, as the Russellian views hold, and that these quiddities along with structure ground all conscious experience, as the constitutive views hold.

Is Russellian monism a form of physicalism, dualism, or something else? As before, this is a largely verbal question that we need not settle. We could say that it is a form of broad physicalism but not narrow physicalism, and leave it at that. Still, it is interesting to look more closely at the question of whether, on a Russellian monist view, (proto)phenomenal properties (that is, phenomenal or protophenomenal properties) are physical properties. There are a number of different options available here, depending on what one counts as a physical property, and how one construes the semantics of physical terms such as ‘mass’. Each of these options leads to a subtly different way of characterizing Russellian monism. The following discussion may be of most interest to aficionados of this topic; others can skip it without much loss.

An initial question is whether physical properties are restricted to the properties invoked by physical theory (space, time, mass, charge, and so on), perhaps along with those properties grounded in them. These are the properties that Stoljar calls the t-physical properties (for theory-physical) and that Strawson (2006) calls ‘physics-al’ properties. It is most common to restrict physical properties in this sense, but one can also invoke expanded senses of the term, such as my notion of a broadly physical property, or Stoljar’s notion of an o-physical property, or Strawson’s notion of a physical property which appears to subsume all natural properties. Given

such an expanded sense, then even if quiddities are not t-physical properties, they may count as physical in the expanded sense. The resulting position might be seen as *expansionary Russellian physicalism*, with (proto)phenomenal properties counting as physical properties in an expanded sense.<sup>11</sup>

In what follows, I will make the more common assumption that physical properties are restricted to t-physical properties: perhaps space, time, mass, charge, and so on. To assess the status of Russellian monism, we can then ask: what is the relationship between (proto)phenomenal properties and physical properties such as mass? This depends on just how terms such as ‘mass’ function.

On one view, ‘mass’ refers to the property that actually plays the mass role. So insofar as there is a quiddity that actually plays the mass role, that quiddity is identical to mass. The corresponding version of Russellian monism is the *Russellian identity theory*, because it holds that (proto)phenomenal properties are identical to physical properties such as mass. As Grover Maxwell (1979) observes, this is a sort of inversion of the more familiar identity theory due to Smart (1961) and others. The familiar identity theory offers a topic-neutral analysis of mental expressions, where ‘pain’ refers to whatever plays the pain role, and then holds that these have physical referents, with C-fiber firing playing the pain role. The Russellian identity theory instead offers a topic-neutral analysis of physical expressions, where ‘mass’ refers to whatever plays the mass role, and then holds that these have mental or proto-mental referents, with (proto)phenomenal quiddities playing the mass role.<sup>12</sup>

On another view, ‘mass’ refers to the second-order functional property of having a property that plays the mass role. On this view, mass is not identical to the quiddity that plays the mass role, but we might say that mass is realized by that quiddity. A closely related view holds that ‘mass’ refers to a dispositional property which is realized by the quiddity that serves as its

---

<sup>11</sup> Stoljar and Strawson are naturally counted as expansionary Russellian physicalists. Strawson spends some time arguing with people like me (for example, denying that physical duplicates without consciousness are conceivable), but once it is clear that I mean by ‘physical’ what Strawson means by ‘physics-al’, the disagreement between us becomes largely verbal.

<sup>12</sup> Maxwell (1979) and Lockwood (1989) are certainly Russellian identity theorists: both explicitly endorse the identity theory and credit the underlying idea to Russell. Feigl (1957) and Montero (2010) can easily be interpreted as holding the view. The coherence of the Russellian identity theory, on which quiddities are identical to t-physical properties and on which t-physicalism is true, suggests that Russellian monism is not best characterized (following Stoljar) as o-physicalism about consciousness without t-physicalism.

categorical basis. The corresponding version of Russellian monism is the *Russellian realization theory*, since it holds that physical properties such as mass are realized by (proto)phenomenal properties. Russellian realization theory can be seen as an inversion of the familiar functionalist realization theory, on which mental properties are second-order functional properties (pain is the property of having a property that plays the pain role) and on which these properties are realized by physical properties.

On the Russellian realization theory, quiddities are not themselves t-physical properties (at least if we assume that realizing properties are distinct from the properties they realize). So the Russellian realization theory is not a version of physicalism, assuming as above that only t-physical properties are physical properties. Instead, physical properties are themselves realized by and grounded in the (proto)phenomenal properties that serve as quiddities. The panpsychist version of this view can be seen as a form of *Russellian idealism*, with fundamental phenomenal properties serving as the grounds for physical properties. The panprotopsychist version can be seen as a form of *Russellian neutral monism*, with fundamental protophenomenal properties serving as the grounds for both physical and phenomenal properties. There may also be a mixed view, perhaps *Russellian pluralism*, if some quiddities are phenomenal and some are protophenomenal or unrelated to the phenomenal.<sup>13</sup>

There are also views on which ‘mass’ on which refers to a dispositional property that is not grounded in its categorical basis: instead categorical and dispositional properties are equally fundamental, and neither is grounded in the other. Given that physical properties are restricted to t-physical properties and those grounded in them, the corresponding version of Russellian monism will be a *Russellian property dualism*, with fundamental physical properties (dispositional properties such as mass) and equally fundamental phenomenal or protophenomenal properties (the corresponding quiddities).

On a final view (the “powerful quality” view advocated by Heil 2012), dispositional properties are identical to their categorical bases. Any corresponding version of Russellian monism will be a version of the Russellian identity theory: whether ‘mass’ functions to pick out a dispositional property or its categorical basis, it will pick out a (proto)phenomenal property. One version of this Russellian identity theory (advocated by Hedda Hassel in forthcoming work) holds that there is a sort of conceptual or a priori connection between (proto)phenomenal

---

<sup>13</sup> Bolender (2001) puts forward a sort of Russellian idealism, and Rosenberg may be either a Russellian idealist or pluralist.

properties and the associated dispositions, in the same way that there is arguably such a connection between pain and certain associated dispositions (arguably, one cannot conceive of pain that does not play a certain dispositional role). Another version, which stands to the first version roughly as type-B materialism stands to type-A materialism, holds that there is an a posteriori connection between (proto)phenomenal and dispositional properties. Note that these versions of the Russellian identity theory are consistent with the version discussed a few paragraphs above, on which (for example) ‘mass’ is equivalent to ‘whatever plays the mass role’. They do not entail it, however, as they are also consistent with views on which ‘mass’ picks out a disposition directly, and they are not entailed by it, as the original version is consistent with views on which dispositional and categorical properties are distinct. One could also see these views as versions of Russellian idealism or neutral monism, on which all truths are grounded in (proto)phenomenal truths.

A number of these versions of Russellian monism differ only verbally. Many of these differences turn on the correct semantics for ‘mass’ and for ‘physical property’, with the underlying metaphysical picture looks the same. One exception here is the difference between Russellian idealism, neutral monism, and pluralism: this turns on the (presumably substantive) issue of whether there is something it is like to have a quiddity. Another may be the differences involving Russellian property dualism and the versions of the Russellian identity theory in the previous paragraph: these turn on the (arguably substantive) issue of whether dispositional properties are grounded in, identical to, or independent of their categorical bases. For what it is worth, I am most attracted to the first version of the Russellian identity theory, with some sympathy also for the idealist, neutral monist, and property dualist versions. The only view that I am entirely unsympathetic with is the a posteriori version of the Russellian identity theory in the previous paragraph (which I think requires a sort of brute identity claim, and so stands to the first version as type-B versions of the familiar identity theory stand to type-A versions). In what follows, I will simply talk of Russellian monism, distinguishing panpsychist and panprotopsychist views as necessary.<sup>14</sup>

---

<sup>14</sup> The different versions of Russellian monism will take different attitudes to the conceivability and possibility of zombies: physical duplicates without phenomenal states. Expansionary Russellian physicalism will deny that they are conceivable or possible: given the expanded sense of the physical, to conceive of a zombie requires conceiving of a categorical zombie (same structure, same quiddities, no consciousness), which cannot be done according to the view. Russellian identity theorists of the first sort discussed above

## 6 Antithesis: The Combination Problem

Given Russellian monism as our new synthesis, a more significant antithesis now threatens. This antithesis takes the form of a major problem for both panpsychism and panprotopsychism: the combination problem.

The combination problem for panpsychism was posed by William James (1890) and named by William Seager (1995). This problem can be stated as follows: how do microexperiences combine to yield macroexperiences? It is at least very hard to see how a number of separate experiences had by separate entities could combine to yield a distinct experience had by a composite entity. It is especially hard to see how they could combine to yield the distinctive kind of macroexperience that we find in our own case.

One way to pose the combination problem is in the form of a conceivability argument. (An approach along these lines is presented by Goff (2009), to whom my presentation here is indebted.) Here  $PP$  is the conjunction of all microphysical and microphenomenal truths about the world, and  $Q$  is a macrophenomenal truth, such as ‘Some macroscopic entity is conscious’.

- (1)  $PP \& \sim Q$  is conceivable.
  - (2) If  $PP \& \sim Q$  is conceivable, it is possible.
  - (3) If  $PP \& \sim Q$  is metaphysically possible, constitutive panpsychism is false.
- 
- (4) Constitutive panpsychism is false.

Here premises (2) and (3) parallel the corresponding premises of the conceivability argument against materialism, and are supported by the same reasons. So the key premise here is premise (1). This premise asserts the conceivability of *panpsychist zombies*: beings that are physically and microphenomenally identical to us (and indeed whole worlds that are physically and microphenomenally identical to ours), without any macrophenomenal states.

---

may hold that zombies are conceivable but not possible, because of nontrivial two-dimensional structure in physical terms (the primary intension of ‘mass’ picks out whatever plays the mass role, the secondary intension picks out the quiddity that actually plays the mass role). Russellian idealists, neutral monists, and property dualists may well hold that zombies are conceivable and possible, in that there are conceivable and possible situations where the structural properties are associated with different quiddities that are independent of the phenomenal, or perhaps with no quiddities at all.

Why believe that panpsychist zombies are conceivable? Some might find this simply intuitive: one can conceive of all the microexperiences one likes without any macroexperiences. But one can also justify it by invoking a principle in the spirit of James' objection to panpsychism in *The Principles of Psychology*. This is the principle that no set of conscious subjects necessitates the existence of a further conscious subject. Or in the key of conceivability: given any any set of conscious subjects and any conscious subject not in that set, one can always conceive of all the subjects in the set without the further subject. More precisely: given any conjunction  $S$  of positive phenomenal truths about a group of conscious subjects and any positive phenomenal truth  $T$  about a conscious subject not in that group,  $S \& \sim T$  is conceivable.

We might say that these principles invoke a subject/subject gap: an epistemic gap from the existence of subjects to the existence of distinct subjects. The principles all have intuitive appeal. *Prima facie*, it seems conceivable that any group of conscious subjects could exist alone, without any further subjects. But if this is right, constitutive panpsychism is in trouble. Given that all experiences are had by conscious subjects, we can say that microexperiences will be had by microsubjects and macroexperiences by macrosubjects. Then by the principle above, we can conceive of any number of microsubjects having their microexperiences without any macrosubject having macroexperiences. That is, we can conceive of the conjunction of all microphenomenal truths obtaining without any positive macrophenomenal truths obtaining.

This result (along with the conceivability-possibility premise) already rules out a version of constitutive panpsychism on which macroexperience is wholly grounded in microexperience. To rule out all versions, including those in which macroexperience is grounded in microexperience plus physical structure, we can appeal to a modified principle according to which in the case above,  $S \& S' \& \sim T$  is conceivable, where  $S'$  characterizes the physical and structural properties of the members of the original group. This principle seems just about as intuitively plausible as the original principle. Given this principle, premise (1) above follows, and if premises (2) and (3) are granted, constitutive panpsychism is ruled out.

One might think that this problem for panpsychism makes things better for panprotopsychism, as panprotopsychism does not need subjects at the bottom level. Nevertheless, there is also a combination problem for panprotopsychism. This is the problem of how protoexperiences can combine to yield experiences.

As with the combination problem for panpsychism, the combination problem for panprotopsychism can be posed in the form of a conceivability argument. Here *PPP* is the

conjunction of all microphysical and protophenomenal truths (or better, purportedly protophenomenal truths, as the combination problem can be used to question whether purportedly protophenomenal properties are truly protophenomenal), and  $Q$  is a macrophenomenal truth, such as ‘Some macroscopic entity is conscious’.

(1)  $PPP \& \sim Q$  is conceivable.

(2) If  $PPP \& \sim Q$  is conceivable, it is possible.

(3) If  $PPP \& \sim Q$  is metaphysically possible, constitutive panprotopsychism is false.

---

(4) Constitutive panprotopsychism is false.

Once again, the key premise is premise (1). This asserts the existence of *protophenomenal zombies*: beings that share our (purportedly) protophenomenal properties at the microphysical level, but without consciousness. The conceivability of protophenomenal zombies is perhaps somewhat less obvious than the conceivability of panpsychist zombies, as we have a less clear idea of what protophenomenal properties involve. Still, one might appeal to a general nonphenomenal/phenomenal gap, as on a view I discussed in section 4. One thought here is that for any nonphenomenal truths, we can conceive of all of these truths obtaining without any experience at all.

Why accept this? One possible justification is a nonsubject/subject gap. This is the claim that no set of truths about nonsubjects of consciousness can necessitate the existence of distinct subjects of consciousness. Or in the key of conceivability: for any set of nonsubjects instantiating nonphenomenal properties and any independent subject exhibiting phenomenal properties, we can conceive of the former without the latter. This principle leads naturally to premise (1) above.

Why believe this principle? One potential justification is the idea that subjects are conceptually fundamental entities. On a view where subjects are metaphysically fundamental entities, then they are not grounded in more fundamental entities, and one can make a case that they are not necessitated by the existence of other fundamental entities. Likewise, if they are conceptually fundamental entities, they are not conceptually grounded in more fundamental entities, and one can make a case that their existence is not a priori entailed by that of other entities. Certainly these principles are not obvious, but they have a certain intuitive appeal.

Another potential justification is a nonquality/quality gap. Here the idea is that phenomenal properties are qualitative, in that they constitutively involve qualities such as redness, greenness, and so on. And one can argue that nonqualitative truths never necessitate qualitative truths, insofar as one can always conceive of the former obtaining without the latter obtaining. Insofar as purportedly protophenomenal properties are nonqualitative, this principle yields a gap between these properties and the phenomenal that can might justify premise (1).

Both panpsychism and panprotopsyichism face challenging combination problems, then. As well as sharing a number of problems, each view faces one especially difficult problem that the other does not: the subject-subject gap for panpsychism, and the nonphenomenal-phenomenal gap for panprotopsyichism. Reasonable people can differ on which problem is more serious. I am inclined to think the subject-subject problem is more difficult, and that panprotopsyichism may benefit from having fewer constraints on its building blocks, but I am far from certain about this. All of these problems have the status of challenges rather than refutations, but they are challenges that need to be addressed.

Of course physicalism is faced with its own version of the combination problem: how do microphysical entities and properties come together to yield subjects, qualities, and so on? This challenge is presumably at least as hard as the challenge to panpsychism, as the resources available to the physicalist are a subset of those available to the panpsychist. But we should be clear on the dialectic. The sympathizer with panpsychism has typically already rejected physicalism (at least in non-Russellian forms), precisely on the grounds of these gaps between the physical and the experiential. The question is then whether panpsychism can do any better. It promises to do better in at least one respect: it accommodates the very existence of experience, if only by taking it as fundamental. But it is not clear whether it does any better at explaining the complex manifest character of macroexperience. This is the challenge posed by the combination problem.

By contrast, dualism does not suffer nearly as badly from a combination problem. This is especially clear for substance dualism, which postulates fundamental entities (subjects of experience) that bear macrophenomenal properties. There is no analog of the subject combination problem for such a view. If the dualist takes macrophenomenal properties as fundamental properties, with their structure, qualities, and other features built in, then there will be no analog of the other combination problems either.



Instead of the combination problem, dualism has the familiar problem of mental causation, as well as a problem of economy (why postulate so many fundamental entities? ). Panpsychism and panprotopsychism, at least in their constitutive Russellian varieties, do not suffer from these problems. They postulate only as many fundamental entities and properties as are needed to make sense of physics (at least if one thinks that physics requires quiddities), and they make a specific hypothesis about the nature of these properties. And on this picture, phenomenal properties are integrated into the causal order.

I think that substance dualism (in its epiphenomenalist and interactionist forms) and Russellian monism (in its panpsychist and panprotopsychist forms) are the two serious contenders in the metaphysics of consciousness, at least once one has given up on standard physicalism. (I divide my own credence fairly equally between them.) So in a way, our new dialectical situation confronts Russellian monism with (once again) substance dualism. In effect the problems of economy and mental causation for one are weighed against the combination problem for the other. If one of these problems can be solved or proved unsolvable, that will constitute significant progress on the mind–body problem.

## **7 New Synthesis: Panqualityism?**

Is a new synthesis in sight? I do not have a solution to the combination problem, so I do not really have a new synthesis. But in this section I want to at least canvas options and to explore one possible new solution, before concluding that it fails. I explore options for dealing with the various aspects of the problem in much more depth in “The Combination Problem for Panpsychism”.

One reaction to the combination problem is to give up on constitutive panpsychism (or panprotopsychism), and instead opt for emergent panpsychism. This view does not face nearly such a pressing form of the combination problem, as it denies that macroexperience is grounded in microexperience. Still, emergent panpsychism loses many of the key advantages of constitutive panpsychism in avoiding the Hegelian dilemma. In particular, it faces a problem of mental causation—how can macroexperience play a causal role? —that is analogous to the problems of dualism, and seems to require epiphenomenalism, interactionism, or overdetermination. So it is worth looking closely at the options for constitutive panpsychism.

A second reaction is to hold that macrosubjects are identical to certain microsubjects: that is, they are identical to certain fundamental physical entities with fundamental phenomenal properties, and they share those phenomenal properties. This view avoids the needs for subjects to combine into distinct subjects. One version of this view is akin to Leibniz's "dominant monad" view, on which human subjects are identical to single fundamental particles, perhaps in their brain. This view is subject to obvious objections, however (what happens if that particle is destroyed? how could a particle have such complex phenomenal properties, especially on a Russellian view? ). Another version of the view appeals to fundamental physical entities above the level of the particle: perhaps entangled quantum systems, or perhaps the entire universe. I think that these possibilities (especially the quantum version) are worth exploring, but it is not easy to see how such entities could have fundamental phenomenal properties that yield a phenomenology like ours.

A third reaction is to *deflate the subject*, either denying that experiences must have subjects at all, or at least denying that subjects are metaphysically and conceptually simple entities. I think it is a conceptual truth that experiences have subjects: phenomenal properties must be instantiated by something, and they characterize what it is like to be that thing. But the second denial seems more tenable. Indeed, some such denial seems required to be a constitutive panpsychist, a constitutive panprotopsychist, or indeed a materialist. This view may require rejecting certain intuitions about subjects, but these intuitions are not non-negotiable.

We might define Subjects as primitive subjects of experience. I think that we have a natural conception of Subjects: these are subjects as they might have been in the Garden of Eden, as it were. I think that where Subjects are concerned, the subject/subject gap and the nonsubject/subject gap are both extremely plausible: the existence of a Subject is not necessitated or a priori entailed by the existence of distinct Subjects or indeed by the existence of non-Subjects. So if we are Subjects (and if we set aside the view that macrosubjects are identical to microsubjects), constitutive panpsychism and constitutive panprotopsychism are false.

Still, it is far from obvious that we are Subjects. There does not seem to be an introspective datum that we are Subjects, and it is not obvious that there are strong theoretical arguments to that effect. There are perhaps intuitions of determinacy about personal identity that tend to support the claim (see Barnett 2010 and Nida-Rümelin 2010), but these intuitions do not seem to be non-negotiable, and there are reasonably strong considerations in favor of rejecting them (see

Parfit 1984). And once we deny that we are Subjects, the door is at least opened to rejecting the subject/subject gap and the nonsubject/subject gap, and to accepting constitutive panpsychism or panprotopsychism.

I think that a Russellian monist must almost certainly embrace this view (perhaps the only remotely promising alternative is the quantum version of the micro/macro identity claim above). Still, to deny that we are Subjects is not to solve the combination problem. We still need to give an account of how macroexperience can be grounded in microexperience or in protoexperience.

Here I will look briefly at a view that has been popular among sympathizers with panpsychism and panprotopsychism: panqualityism. The name of this view was introduced in an article by Herbert Feigl (1960), who credits the term to conversation with Stephen C. Pepper, but versions of view itself were popular among the neutral monists of the early twentieth century, including William James (1904), Ernst Mach (1886), and Bertrand Russell (1921). More recently, the view has been defended by Sam Coleman (2012).

On this view, *qualities* are the properties presented in experience: Intuitively, these are properties like redness, greenness, heat, and so on. Qualities are not identical with phenomenal properties: when redness is presented to me in experience, I have a phenomenal property, but I need not be red. Instead, we would intuitively say that I am aware of redness, and that phenomenal properties involve awareness of qualitative properties. Likewise, phenomenal properties are always instantiated by conscious subjects, but qualities need not be. We can certainly make sense of the idea of a red object that is not a subject of experience.

Panqualityism typically requires rejecting a reductionist view of qualities, such as a view on which color qualities are identified with physical reflectance properties or something of the sort. Instead, it is naturally associated with what I have called an Edenic view of qualities. Here the qualities most fundamentally presented in experience are properties such as Edenic redness, a simple property that may not be instantiated by the objects that seem to have it in the external world, but which might have been instantiated in the Garden of Eden.

Panqualityism holds that fundamental physical entities instantiate qualities like these. We might imagine, for example, that fundamental particles are Edenically red. More likely, the relevant qualities involved will be more austere than this, but they will nevertheless be primitive properties that could be presented in experience. The most important kind of panqualityism, unsurprisingly, is constitutive Russellian panqualityism, on which qualities serve as quiddities

and also serve to constitute human experience. Many of the panqualityists discussed above have endorsed views of this sort.

Constitutive panqualityism is a form of panprotopsychism rather than panpsychism: qualities are not phenomenal properties but serve to constitute phenomenal properties. Because qualities need not be instantiated by subjects, the view need not invoke microsubjects at all. Panqualityism is occasionally characterized as a version of panpsychism with “experiences without subjects” or “unsensed sensa”, but I think the view is best regarded as a form of panprotopsychism. Still, it is a form on which the protophenomenal properties take an especially familiar form, and on which they have a close connection to phenomenal properties.

Panqualityism is not threatened by the subject/subject gap, as it does not require microsubjects to constitute macrosubjects. Likewise, it is not threatened by the nonquality/quality gap, as the purportedly protophenomenal properties here are qualitative through and through. It is threatened by the nonsubject/subject gap, but here it responds by deflating the subject. Some of the traditional panqualityists rejected subjects of experience altogether, while others have taken deflationary views of them on which they can be constituted by underlying qualities, perhaps along with structural properties.

How does panqualityism solve the combination problem? It is natural for the panqualityist to argue that simple microqualities can collectively constitute complex macroqualities, ultimately building up something as complex as the qualitative structure of a visual field, or even a full multisensory field. Then it could be suggested that the existence of these complex qualities explains the phenomenal data even without postulating an associated subject of experience; or it could be suggested that certain complex qualities entail the existence of an associated subject, perhaps in a deflated sense.

Still, I think that panqualityism is vulnerable to a version of the combination problem analogous to earlier versions. In particular, we can mount a conceivability argument against panqualityism as follows. Here  $QQ$  is a conjunction of positive qualitative truths at the microphysical, perhaps along with any other microphysical truths, and  $Q$  is a positive macrophenomenal truth.

- (1)  $QQ \& \sim Q$  is conceivable.
- (2) If  $QQ \& \sim Q$  is conceivable, it is metaphysically possible.
- (3) If  $QQ \& \sim Q$  is metaphysically possible, constitutive panqualityism is false.

(4) Constitutive panqualityism is false.

Again, all the action is in the first premise. This premise asserts the conceivability of *qualitative zombies*, beings that are qualitatively (and microphysically) identical to us without consciousness.

Why believe this premise? One could make a case that it is intuitively obvious. But more deeply, it is grounded in what we might call the quality/awareness gap. Here the idea is that no instantiations of qualities ever necessitate awareness of qualities. Or in the key of conceivability: for any set of instantiated qualities and physical properties, it is conceivable that all those qualities and properties are instantiated without any awareness of the qualities. Given that all phenomenal properties involve awareness of qualities, premise (1) above follows. And even if only some phenomenal properties involve awareness of qualities, this will be enough to make the case against constitutive panqualityism.

The quality/awareness gap has much intuitive force. On the face of it, it is conceivable that Edenic redness is instantiated without anyone being aware of it. And on the face of it, this intuition scales up to arbitrarily complex qualities. Even given complex qualities corresponding to the structure of a visual field, then if it is conceivable that those qualities be instantiated at all (presumably by a situation in the world corresponding to the situation as perceived), it is conceivable that they be instantiated without any awareness of those qualities.

The panqualityist might respond in various ways. They could bring in awareness at the fundamental level, perhaps by appealing to special qualities that cannot be instantiated without awareness of those qualities (pain, perhaps? ); but this leads back to subjects at the fundamental level, and the associated problems. They might deny the existence of awareness, as James (1904) does, and hold that our experience involves qualities but does not involve awareness of them; but this claim runs directly counter to our phenomenology. They might combine the appeal to qualities with a functional reduction of awareness, as Coleman (2012) does; but I think that the conceivability argument above itself gives reason to reject such a reduction.

Panqualityism is also vulnerable to other aspects of the combination problem. It is vulnerable to the *structure combination problem*: the structure among qualities instantiated in the brain is very different from the structure among qualities of which we are aware, and it is hard to see how the former could constitute the latter. It is also vulnerable to the *quality combination problem*: it is hard to see how a few primitive qualities (which is all that the

Russellian panqualityist can appeal to) could yield the vast array of qualities of which we are aware.

I conclude that panqualityism does not offer a solution to the combination problem. We are still in need of a new synthesis.

## 8 Conclusion

We started with the thesis of materialism and the antithesis of dualism, and reached the synthesis of panpsychism. This synthesis encountered the antithesis of panprotopsychism, from which we reached the new synthesis of Russellian monism. This synthesis encountered the antithesis of the combination problem, and whether there can be a new synthesis remains an open question.

Still, I think that the Hegelian argument gives good reason to take both panpsychism and panprotopsychism very seriously. If we can find a reasonable solution to the combination problem for either, this view would immediately become the most promising solution to the mind–body problem. So the combination problem deserves serious and sustained attention.

## References

- Barnett, D. 2010. You are simple. In (G. Bealer and R. Koons, eds.) *The Waning of Materialism*. Oxford University Press.
- Bennett, K. 2003. Why the exclusion problem seems intractable and how, just maybe, to tract it. *Noûs* 37:471-97.
- Bolender, J. 2001. An argument for idealism. *Journal of Consciousness Studies* 8(4): 37-61.
- Chalmers, D.J. 1997. Moving forward on the problem of consciousness. *Journal of Consciousness Studies*.
- Chalmers, D.J. 1999. *Precis of The Conscious Mind. Philosophy and Phenomenological Research*.
- Chalmers, D.J. 2003. Consciousness and its place in nature. In (S. Stich and T. Warfield, eds.) *Blackwell Guide to the Philosophy of Mind*. Blackwell.
- Chalmers, D.J. 2010. The two-dimensional argument against materialism. In *The Character of Consciousness*. Oxford University Press.
- Chalmers, D.J. 2012. *Constructing the World*. Oxford University Press.

Chalmers, D.J. forthcoming. The combination problem for panpsychism. In (G. Bruntrup and L. Jaskolla, eds.) *Panpsychism*. Oxford University Press.

Coleman, S. 2012. Mental chemistry: Combination for panpsychists. *Dialectica* 66:137-66.

Feigl, H. 1958/1967. The 'mental' and the 'physical'. *Minnesota Studies in the Philosophy of Science* 2:370-497. Reprinted (with a postscript) as *The 'Mental' and the 'Physical'*. University of Minnesota Press.

Feigl, H. 1960. Mind-body, not a pseudo-problem. In S. Hook (ed.), *Dimensions of Mind*. New York University Press

Fine, K. 2012. The pure logic of ground. *Review of Symbolic Logic* 5:1-25.

Goff, P. 2009. Why panpsychism doesn't help to explain consciousness.

Heil, J. 2012. *The Universe As We Find It*. Oxford University Press.

James, W. 1890. *The Principles of Psychology*. Henry Holt. *Dialectica* 63:289-311.

James, W. 1904. Does "consciousness" exist? *Journal of Philosophy, Psychology, and Scientific Methods* 1:477-491.

Lockwood, M. 1989. *Mind, Brain, and the Quantum*. Oxford University Press.

Lockwood, M. 1993. The grain problem. In (H. Robinson, ed) *Objections to Physicalism*. Oxford University Press.

Mach, E. 1886. *The Analysis of Sensations and the Relation of Physical to the Psychical*. Translated by C. M. Williams, Open Court, 1984.

Maxwell, G. 1979. Rigid designators and mind-brain identity. *Minnesota Studies in the Philosophy of Science* 9:365-403.

Montero, B. 2010. A Russellian response to the structural argument against physicalism. *Journal of Consciousness Studies* 17:70-83.

Nida-Rümelin, M. 2010. An argument from transtemporal identity for subject-body dualism. In (G. Bealer and R. Koons, eds.) *The Waning of Materialism*. Oxford University Press.

Papineau, D. 2002. *Thinking About Consciousness*. Oxford University Press.

Parfit, D. 1984. *Reasons and Persons*. Oxford University Press.

Russell, M. 1921. *The Analysis of Mind*. George Allen and Unwin

Russell, B. 1927a. *An Outline of Philosophy*. George Allen and Unwin.

Russell, B. 1927b. *The Analysis of Matter*. Kegan Paul.

Schaffer, J. 2009. On what grounds what. In (D. Manley, D.J. Chalmers, & R. Wasserman, eds.) *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University Press.

Smart, J.J.C. 1959. Sensations and brain processes. *Philosophical Review* 68:141-56.

Stoljar, D. 2001a. The conceivability argument and two conceptions of the physical. *Philosophical Perspectives* 15:393-413.

Stoljar, D. 2001b. Two conceptions of the physical. *Philosophy and Phenomenological Research* 62:253-81.

Stoljar, D. 2013. Four kinds of Russellian monism. In (U. Kriegel, ed.) *Current Controversies in the Philosophy of Mind*. Routledge.

Strawson, G. 2006. Realistic monism: Why physicalism entails panpsychism. *Journal of Consciousness Studies*.

Yablo, S. 1992. Mental causation. *Philosophical Review* 101:245-280.