Consciousness and Categorical Properties

Christopher Devlin Brown¹

Received: 20 September 2019 / Accepted: 16 December 2020

© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract

Russellian physicalism is a view on the nature of consciousness which promises to satisfy the demands of both traditional physicalists and non-physicalists. It does so by identifying subjective experience with physically acceptable categorical properties underlying structural and dispositional properties described by science. Though promising, the view faces at least two serious challenges: (i) it has been argued that science deals in both categorical and non-categorical properties, which would undercut the motivation behind Russellian physicalism, and (ii) it has been argued that only nonphysicalist Russellian views—like panpsychism—are useful when it comes to explaining consciousness. I address these criticisms, arguing that there is no viable reason for maintaining that science deals in categorical properties of the sort which a Russellian physicalist is interested in, and that features of fully-physical categorical properties can be described which provide useful explanations for various essential features of subjective experience. These projects are connected: it turns out that when the explanatory relevance of Russellian physicalism is explained in detail, constraints are put the sort of categorical properties that Russellian physicalists can say are left out of science. Specifically, Russellian physicalists are forced to subscribe to the view that science leaves out any categorical properties whatsoever, as opposed to the view that some scientifically scrutable properties are categorical, but not the ones which Russellian physicalists are interested in. I hope that by addressing these criticisms of Russellian physicalism, and drawing logical connections between the responses, further appeal is added to a promising, but so far relatively unexplored, view.

The hard problem of consciousness (Chalmers 1995) emerges from two seemingly inconsistent claims.

Much thanks to Torin Alter and to anonymous Erkenntnis referees for useful feedback on prior drafts.

✉ Christopher Devlin Brown
christopher.devlin.brown@gmail.com

¹ Bilkent University, Ankara, Turkey

Inscrutability subjective experience is not reducible to, explainable by, nor inferable from scientifically scrutable properties

Physicalism all properties are nothing over and above physical properties.

Many philosophers find each of these claims appealing: the arguments of Kripke (1980), Nagel (1974), Jackson (1982,1986), Levine (1983), Chalmers (1996,2010) and others support the intuitive view that *something significant is left out of the scientific account of the mind*; namely, subjective experience.¹ According to these thinkers, no amount of scientific knowledge can possibly tell you *what it is like* to have some experience. The only way to acquire knowledge of some subjective experience is to have the experience. Similarly, many philosophers are attracted to physicalism. They think that this is the proper articulation of a metaphysics of science, and it is supported by the causal closure argument (Papineau 2001), which concludes that all physical effects—including human behavior—have only physical causes.

These two views seem inconsistent because many assume that if some property is physical, then it is scientifically scrutable. Unfortunately, both “scientifically scrutable” and “physical” are imprecise. For present purposes, “scientifically scrutable” can be understood as the epistemic condition of being knowable through scientific inquiry, and “physical” can be understood as the metaphysical condition of not being identical to, dependent upon, or constituted by anything which is fundamentally mental.² Something is fundamental if it depends on nothing else; for example, if biological and chemical and atomic properties ultimately depend on subatomic properties, and subatomic properties depend on nothing else, then subatomic properties are fundamental.³ A property which seems to be both scientifically scrutable and physical is solidity, which is scientifically explained by the molecular and lower-level properties of a solid object, and which seems not to depend on any properties which are fundamentally mental (putting aside panpsychism, for the moment). Conversely, souls seem to be both scientifically inscrutable and nonphysical: they are

¹ One might think that some of these arguments straightforwardly contradict physicalism, rather than supporting inscrutability. For instance, the conceivability argument (Chalmers 1996) typically concludes that physicalism is false. However, as Chalmers (2015) acknowledges, this argument tacitly assumes that physical properties are scientifically scrutable properties, and the argument fails to rule out physicalism if a broader view of physical properties is assumed.

² This definition of ‘physical’ is consistent with *via negativa* physicalism (Smith 1993; Papineau and Spurrett 1999; Levine 2001; Montero 2005; Montero and Papineau 2005; Worley 2006), which says the essence of a physical property is that it is not identical to nor dependent upon anything which is fundamentally mental, but not theory-based physicalism (Hellman and Thompson 1975; Smart 1978; Hellman 1985; Poland 1994; Melnyk 1997, 2002; Stoljar 2001), which says that ‘physical’ means ‘a posit of physics, or entirely dependent upon posits of physics’. For present purposes it suffices to take the *via negativa* understanding of physical as a sufficient-for-false condition on physicalism, such that the existence of fundamentally mental properties would make physicalism false. Nearly everyone accepts this sufficient-for-false condition, even if they think a full understanding of the physical requires something else in addition (e.g. Wilson 2006).

³ Whether fundamental properties must be properties of fundamental objects is a metaphysical issue that Russellian physicalists can remain neutral on. As an alternative to a substance-based metaphysics, bundle theory says that there are *only* properties which are bundled together, and no additional substances or non-property-objects which the properties are attached to.

fundamentally mental entities which cannot be fully understood through scientific inquiry.

Recently, a view called ‘Russellian physicalism’—a version of a wider view called ‘Russellian monism’—has emerged as a solution to the hard problem of consciousness (see Stoljar 2001; Montero 2010, 2015; Chalmers 2003, 2015; Pereboom 2011, 2013; Alter and Nagasawa 2012; McClelland 2013; Alter 2016; Brown 2017). Russellian physicalism says that a property can be physical without being scientifically scrutable. This is because science seems to only deal with dispositional and structural properties of things, such as their spacetime and causal properties. However, one might think that physical properties need not be constrained to only the dispositional or structural: a physical property might instead be “categorical”, which just means a property that is not entirely dispositional or structural. The categorical properties I’m interested in are (to borrow terminology from Pereboom 2011) absolutely intrinsic: they do not entirely depend on more fundamental properties which are dispositional or structural, though fundamental categorical properties may form the dependence base for higher-level dispositional or structural properties.

Russellian physicalists address the hard problem in the following way: subjective experience is identical to a property which is both physical and categorical. Assuming that science only reveals non-categorical properties, this position is able to respect both the inscrutability of phenomenal properties, as well as physicalism. What distinguishes Russellian physicalism from non-physicalist versions of Russellian monism (e.g. Strawson 2006; Chalmers 2015; Mørch 2018a) is that Russellian physicalism denies that fundamental categorical properties are mental or protomental.⁴ This is in contrast with panpsychism or panprotopsychism, which respectively hold that fundamental entities like quarks (assuming that quarks are fundamental) possess fundamental mentality or protomental. It follows that Russellian physicalists are committed to the existence of non-fundamental categorical properties (which themselves ultimately depend on fundamental non-mental categorical properties), and that phenomenal properties are identical to some of these high-level categorical properties.

If successful, Russellian physicalism is a view that offers significant appeal: it promises to satisfy the demands of both traditional physicalists and non-physicalists. Unfortunately, even if one assumes the independent plausibility of physicalism and the inscrutability of subjective experience, Russellian physicalism faces at least two serious challenges: (i) Hiddleston (2019) argues that there is a continuum between fully categorical and fully non-categorical properties, and that science deals with

⁴ By “protomental” I mean a property which is especially suited to generating mentality. These properties are unlike e.g. the subatomic properties as discussed by particle physicists, which may be capable of generating higher-level mentality, but are not *especially* suited to this purpose. A counterfactual test devised by Morris (2016) can determine whether a property is protomental in a physically problematic way: if it is possible for some world’s protomental properties to be swapped with non-protomental properties, and there would be no other change *save a difference in mentality*, then the protomental properties of that world are physically unacceptable. When I speak of protomental properties, I am only thinking of physically unacceptable protomental properties. Further arguments that protomental properties of this sort are unacceptable under physicalism can be found in Howell (2009) and Brown (2017).

properties in the middle of the continuum, which have both categorical and non-categorical aspects; and (ii) one might think that the only versions of Russellian monism which have any hope of addressing the hard problem of consciousness are panpsychism or panprotopsyism, which rules out Russellian physicalism.⁵

Though seemingly independent issues, they are closely related: both criticisms say that Russellian physicalism is undercut once we proceed with a serious analysis of categorical properties. The difference is that these criticisms come at Russellian physicalism from opposite directions. The first says that Russellian physicalists are wrong to assume that categorical properties are not scientifically scrutable. If science does indeed traffic in the sort of categorical properties that define Russellian physicalism, then its motivation is undercut, since these categorical properties would be in the set of scientifically scrutable properties that are the target of standard anti-physicalist arguments. The Russellian physicalist would be wrong when she says that our typical conception of the scientifically scrutable leaves out consciousness-relevant categorical properties.

The other criticism assumes the feasibility of a Russellian solution to the hard problem, but denies that Russellian physicalism in particular can provide the solution. This can be put as an abductive argument: so far as we can tell, the only categorical properties capable of explaining human consciousness are fundamentally mental properties such as those posited by panpsychism. No other sort of categorical property we can imagine—solidity, color, “prime materiality”, etcetera—is sufficient to do the job.

My goal here is not just to defend Russellian physicalism from these two criticisms, but to show that the proper responses to these criticisms can help shape the view’s contours, which is relatively unexplored and in need of such shaping. I will proceed as follows: section one will lay out and respond to Hiddleston’s arguments concerning the scientific scrutability of categorical properties. Section two will reconstruct and respond to the abductive argument that panpsychism is the only version of Russellian monism which can address the hard problem. Section three will use the responses from sections one and two as a basis for fleshing out the details of Russellian physicalism.

1 Super-Categorical and Super-Dispositional Properties

Russellian physicalism says that there are role-fillers of fundamental physics which are both categorical and inscrutable. Under this view, an entity like an electron has a variety of dispositional and structural properties which are described by physics, and

⁵ These are not the only criticisms of Russellian physicalism, but I believe other criticisms have been sufficiently dealt with elsewhere. For instance, it has been argued that Russellian physicalism is a disguised version of non-physicalist panprotopsyism (Morris 2016); that it leads to epiphenomenalism (Howell 2015); and that it cannot solve the combination problem of generating human consciousness from non-human-conscious properties (Chalmers 2016). I appeal to the respective solutions to these criticisms from myself (Brown 2017), Alter and Coleman (2019, forthcoming), and McClelland (2013), though versions of these issues will come up as I progress.

additionally has one or more categorical properties underlying its dispositional and structural properties which are not revealed through scientific inquiry. Science can tell us the electron's location in spacetime, how it behaves, and how it interacts with other things. Science, however, does not tell us anything whatsoever about *what the electron intrinsically is* independent of a dispositional and structural characterization. The categorical, according to Russellian physicalism, is left out.⁶

This view—that science omits fundamental categorical properties—is what allows the Russellian physicalist to bypass the traditional consciousness-based arguments against physicalism. Consider the conceivability argument (Kirk 1974; Chalmers 1996, 2010). This argument says that if we can conceive of a minimal physical duplicate of our world—a duplicate of all and only our world's fundamental physical properties, and whatever properties they metaphysically necessitate—which has different phenomenal properties than our world, then such a world is possible. If it is possible that a minimal physical duplicate with different phenomenal properties is possible, then physicalism is false. Such a world is conceivable, so physicalism is false.

Russellian physicalists urge that when we have so-far conceived of minimal physical duplicate worlds, we have been conceiving of non-categorical properties only. This is because we have so-far conflated 'physical' and 'scientifically scrutable', and the set of all scientifically scrutable properties completely leaves out consciousness-relevant categorical properties.⁷ According to Russellian physicalists, a more robust conception of the physical shows that we cannot properly conceive of a minimal physical duplicate of our world which has different phenomenal properties, since phenomenal properties are constituted or determined by fundamental categorical properties. This move is supposed to satisfy the demands of non-physicalists: Russellian physicalists admit that there are absolutely intrinsic, ineffable properties which we have direct access to in subjective experience, and which are completely outside the ken of science.

Under Russellian physicalism, zombie worlds are in a sense conceivable (since there could be worlds which duplicate all and only the scientifically scrutable properties of our world, and which lack any phenomenal properties whatsoever), knowledge of phenomenal properties can only be acquired by having experiences, and

⁶ An alternative and less popular way of articulating Russellian physicalism says that science may invoke categorical properties of some sort, but that there are more fundamental *pure* categorical (or "super-categorical") properties underlying the posits of science. I will address both of these routes to Russellian physicalism.

⁷ It is worth mentioning the "powerful qualities" view, championed by Martin (1997) and Heil (2005). Under this view, properties are simultaneously categorical and dispositional, and neither aspect is prioritized over the other. I must admit that I have some difficulty understanding this position—since "categorical" means "non-dispositional", and it seems somewhat odd that one and the same property can be both dispositional and non-dispositional—but let me nonetheless try to describe the relationship between this view and Russellian physicalism. If all properties can be correctly referred to via a description or concept which invokes only dispositions, then all properties are scientifically scrutable, even if science only employs disposition-invoking descriptions. As far as I can tell, this would render the Russellian physicalist position on the hard problem nonviable, which means that Russellian physicalism and the powerful qualities view are not consistent with each other.

there is an unbridgeable explanatory gap between a scientific understanding of the brain and consciousness—yet physicalism is nonetheless true. Nearly all parties in the debate seem to get what they want (except for those who think that there is no explanatory gap between scientifically scrutable and phenomenal properties, e.g. Dennett 1991).

As you can see, a lot turns on the claim that the set of scientifically scrutable properties excludes the Russellian physicalist's categorical properties. If the set of all scientifically scrutable properties includes the set of all categorical properties, then the conceivability argument still defeats physicalism: it would turn out that we *do* conceive of all the categorical properties of our world when we conceive of a world which duplicates all and only our world's scientifically scrutable properties. So too with the knowledge argument (Jackson 1982, 1986): it would be the case that Mary *already possessed knowledge of all the categorical properties* when she was in the black and white room, yet she nevertheless seems to learn something new when she leaves.

Is there any reason to think that science deals in categorical properties of the sort which Russellian physicalists are interested in? Hiddleston (2019) provides arguments to this effect. He says that Russellian monism (his arguments target both physicalist and non-physicalist versions) requires either that consciousness-relevant categorical properties be super-categorical, or that scientifically scrutable properties be super-dispositional, and that neither of these routes are viable. Allow me to explain.

A super-categorical property is a property which has no essential non-categorical features at all. When I talk about the “features” or “aspects” of a property, I am referring to higher-order properties possessed by the lower-order property; e.g. the property of *being red* has the higher-order property of *being a color*.⁸ Super-categorical properties are in principle capable of playing any dispositional or structural role whatsoever, since whatever role they play in a world is merely nomologically determined, rather than being metaphysically determined. For instance, it is metaphysically contingent that such a super-categorical property plays the mass role in some world rather than the charge role.

On the other end of the spectrum, a super-dispositional property is a property which has no categorical features whatsoever. Such a property is entirely defined by whatever dispositional or structural role it plays. As I understand it, Russellian monists more commonly endorse the view that all scientifically scrutable properties are super-dispositional, though sometimes they also say that the consciousness-relevant categorical properties they are interested in—and which are left out of science—are super-categorical.

⁸ Since I am using the term ‘feature’ to refer to *any* property of a property, it will turn out that non-categorical properties will have features which mention their categorical role-fillers; e.g. non-categorical property N has the feature of depending upon categorical property C. This is not an issue for Russellian monism so long as categorical-mentioning features of non-categorical properties are omitted from science. Similarly, super-categorical properties will have non-categorical properties, e.g. C has the property of forming part of the dependence base for N. This is not an issue for super-categorical Russellian monism so long as these non-categorical features are not essential to super-categorical properties.

As I said, Hiddleston asserts that Russellian monism requires either (i) the existence of consciousness-relevant scientifically inscrutable properties which are super-categorical, or (ii) that all scientifically scrutable properties are super-dispositional. Up to this point, I think he is right. It can only be the case that science leaves out consciousness-relevant categorical properties if those properties have no dispositional or structural aspects whatsoever, since all scientifically scrutable properties seem to have at least some non-categorical aspects; or if scientifically scrutable properties are not at all categorical, since the sorts of properties that Russellian monists are interested in would trivially be left out of science. The other option—that scientifically scrutable properties have both categorical and non-categorical features, and that the properties which a Russellian monist is interested in have both categorical and non-categorical features—is no good: in this case, science would employ properties of exactly the sort that the Russellian monist accuses science of leaving out.

Hiddleston further argues that neither option available to the Russellian monist will work. He provides two arguments: I'll call them the "fried egg" and the "proton/anti-proton" arguments. The fried egg argument targets the existence of super-categorical properties. It says that since super-categorical properties can play any structural or dispositional role whatsoever, it follows that the categorical property underlying *being an electron* could equally well play the role of the categorical property underlying *being a fried egg*, and vice versa. Hiddleston then correctly points out that this is absurd: it seems metaphysically impossible that an electron's categorical role-filler could possibly play the role of a fried egg.

The proton/anti-proton argument is set up as follows: imagine a world in which there are protons and anti-protons. Both sorts of entities have the same dispositional profiles, and they annihilate each other on contact. It nonetheless seems that scientists would be able to intelligibly and justifiably refer to and discuss the two sorts of entities, despite the fact that they are dispositionally identical. Since a categorical property is defined as a property which is not entirely dispositional or structural, it seems to follow that the scientists are referring to categorical properties when they discuss protons and anti-protons. Thus, scientifically scrutable properties, like the property of *being a proton* or *being an anti-proton*, are not super-dispositional. Neither of these arguments ultimately succeeds—I'll address each in turn.

The fried egg argument alleges that the Russellian monist is committed to the view that any categorical property whatsoever can play any role whatsoever, with the consequence that a fried egg can play the role of an electron and vice versa. However, a Russellian monist is not so committed. She can instead restrict super-categorical properties to only *fundamental* categorical properties. On this narrower version of Russellian monism, the super-categorical property that plays the mass role could instead play the charge role (assuming mass and charge are fundamental), but it could not play the non-fundamental role of *being a fried egg*. As I see it, it is far less counterintuitive that whatever plays the mass role could swap with whatever plays the charge role, as opposed to it being able to swap with the fried egg role. This is no different from the well-known view of Lewis (2009) that the world consists of a Humean mosaic of fundamental categorical properties, the distribution of which determines both their nomological role and whatever other non-fundamental

properties there are in existence.⁹ Though non-categorical properties require categorical properties as relata which figure in appropriate relations—and so non-categorical properties partially depend on fundamental categorical properties—*any* fundamental categorical relata will suffice. Under this account, relations between fundamental categorical properties are doing the heavy metaphysical lifting when it comes to determining particular non-categorical properties, rather than the categorical relata.

This response to the fried egg argument works because Russellian monism is not committed to the existence of non-fundamental super-categorical properties. Instead, a Russellian monist must say that phenomenal properties cannot be identical to nor entirely dependent upon non-categorical properties. This condition is satisfied if only some properties (e.g. fundamental properties) are super-categorical, since fundamental super-categorical properties can partially constitute the dependence base for non-fundamental, non-super-categorical properties. Higher-level categorical properties—like human phenomenal properties, or perhaps the categorical property underlying a fried egg—can have essential structural or dispositional features, so long as they depend on some purely categorical properties. These essential dispositional or structural features prevent *all* categoricals from being swapped willy-nilly with *any* other categorical properties.

Though the fried egg argument ultimately fails, it does illustrate something important about “super-categorical Russellian monism” (i.e. Russellian monism which commits to the existence of super-categorical properties that are left out of science): this view requires that there be non-fundamental categorical properties which are not super-categorical, and which at least partially depend on fundamental super-categorical properties. Further, fundamental super-categorical properties are metaphysically capable of swapping with each other.¹⁰ These entailments will be useful when later assessing the best way to articulate Russellian monism.

⁹ This account assumes that dispositional properties depend on spacetime relations between categorical relata, and that any fundamental categorical relata suffice to generate spacetime-dependent dispositions. However, one might wonder why property swaps among fundamental categorical properties are less absurd than property swaps among non-fundamental categorical properties—perhaps the intuition that swaps of the former sort are less absurd than swaps of the latter sort comes from a bias concerning the ontological specialness of middle-level objects like ourselves and the objects we are directly acquainted with in perception. Though this may be, it does not significantly affect my overall argument, since (i) this sort of Humean mosaic account of reality is quite popular, and (ii) I will end up defending super-dispositional Russellian physicalism rather than super-categorical Russellian physicalism.

¹⁰ One might worry that these commitments of super-categorical Russellian monism collectively entail that fundamental categorical properties are protomental in a way which precludes them from being physically acceptable, similar to a problem discussed by Morris (2016) and myself (Brown 2017) for Russellian physicalism generally. This is because it seems that categorical properties make no difference to the properties of physics, since the properties of physics are non-categorical and swapping fundamental super-categorical properties has been stipulated to make a difference to only higher-level *categorical* properties. If all non-fundamental categorical properties are phenomenal properties, then it seems that super-categorical properties are mentality-specific. A super-categorical Russellian physicalist has a way out of this problem, which follows my (Brown 2017) lead: if there are non-fundamental categorical properties which are non-mental, then swapping fundamental super-categorical properties will affect these non-mental properties as well. Assuming that the notion of a non-mental, non-fundamental categorical property is coherent, this allows super-categorical Russellian monists to endorse physicalism.

I've just shown why the fried egg argument doesn't work, and it turns out that the proton/anti-proton argument fails as well. This is because—contra Hiddleston's assessment—the scientists in the proton/anti-proton argument distinguish between protons and anti-protons via their dispositional and structural properties. Specifically, they are using the spacetime and causal properties of protons and anti-protons.

Both protons and anti-protons possess a *nearly* identical dispositional characterization (characterization "D"), with the exception that protons are disposed to mutually annihilate anti-protons and not other protons, and anti-protons are disposed to mutually annihilate protons and not other anti-protons. This difference in disposition is what allows scientists to justifiably posit two sorts of things: they observe that some things that satisfy characterization D mutually annihilate some other things satisfying D, but not everything satisfying D mutually annihilates anything else satisfying D. In this way, the *non-identical* dispositional features of protons and anti-protons provide the basic epistemic justification for positing two sorts of things.

When specifically fixing reference of the kind term "proton", the scientists use the spacetime properties of the entities in the extension set to fix the reference. They say, "*these* things (in such-and-such spacetime region) are protons, and whatever annihilates them are anti-protons". Alternately, the scientists might use "proton" to refer to whichever sort of things they themselves are made of, and "anti-proton" to refer to things which annihilate the protons that they are made of, but these too are spacetime properties—they are using the things that compose the bodies in the spacetime region they themselves occupy in order to fix the reference.

Two things are going on here: the scientists are positing two sorts of things in an epistemically justified way, and they are successfully fixing natural kind terms to the two sorts (both tasks would likely be accomplished simultaneously). Neither procedure requires invoking categorical properties. The scientists use observed dispositional properties as their epistemic grounds for positing two classes of entities, and they fix references to entities in the two classes via the entities' structural properties. Thus, scientists can successfully posit and refer to protons and anti-protons without ever employing categorical properties, and this scenario does not present a counterexample to the view that all scientifically scrutable properties are super-dispositional. Note that, unlike the response to the fried egg argument, this response to the proton/anti-proton argument does not put constraints on how Russellian monism must be characterized. Thus, "super-dispositional Russellian monism"—which says that scientifically scrutable properties are super-dispositional—has fewer metaphysical requirements than super-categorical Russellian monism.

One might object that, though the scientists employ only non-categorical properties when they posit and fix references to protons and anti-protons, they nonetheless indirectly refer to the categorical role-fillers underlying those non-categorical properties. This is similar to referring to a specific person via a description of her non-essential properties, e.g. I might refer to someone named "Devin" via the property of being *the person wearing a red hat at such-and-such time and place*. The issue with this response is that such indirectly-referred-to properties seem not to be part of science. Suppose that panpsychism is true. In this case, fundamental entities are essentially mental, and science refers to their mentality, *yet this mentality is not an explicit part of any scientific theory whatsoever*. If fairies were the role-fillers

of physics (as Hiddleston facetiously suggests), then the same theory would refer to fairies. Even if there is a sense in which science indirectly refers to categorical properties, categoricals don't count as scientifically scrutable unless they are directly referred to.

Though Hiddleston is right that Russellian monists require either the existence of super-categorical properties or that all scientifically scrutable properties are super-dispositional, he is wrong that neither of these options are viable. However, there remains a question—which I will address in section three—about which of these two options is preferable for a Russellian physicalist.

2 The Abductive Argument for Panpsychism

So far, I've defended the metaphysical assumptions required by any version of Russellian monism. Now I will specifically defend Russellian *physicalism* from an argument that it cannot solve the hard problem, whereas panpsychism can. I'll call this the "panpsychic challenge" (it is similar to arguments articulated in Seager 2006; Strawson 2006; Goff 2009; Alter and Nagasawa 2012; Chalmers 2015; Mørch 2018a).

1. For any type of categorical property *C*, *C* suffices to solve the hard problem only if *C* can explain human phenomenal properties
2. Fundamentally mental or fundamentally protomenta properties are the only types of categorical properties that can explain human phenomenal properties
3. Russellian physicalism posits neither fundamentally mental nor fundamentally protomenta properties
4. Therefore, Russellian physicalism cannot solve the hard problem

The idea is this: imagine any sort of non-mental (and non-protomenta) fundamental categorical property you wish. Some proposed examples (all well discussed by Pereboom 2011—my initial analysis is indebted to his) are categorical solidity (advocated by Locke and Newton), prime materiality (advocated by Aristotelians), and primitive color (Armstrong 1961; and similar to Coleman 2015, 2016). None of these seem to be capable of explaining human consciousness. In fact, the only properties that seem to be capable of serving that explanatory role are properties that are themselves phenomenal, or somehow closely tied to phenomenology. However, no such fundamentally mental or protomenta property is acceptable under Russellian physicalism. Since Russellian physicalists cannot appeal to any sort of fundamental categorical property that would plausibly explain human consciousness, they fail to solve the hard problem.

Categorical solidity, prime materiality and primitive color are exemplars of three routes toward fleshing out the details of the nature of fundamental categorical properties. Each proposal has a problem which renders it nonviable. Specifically: (i) if fundamental categorical properties are very closely related to dispositional or structural properties, then they are unable to explain essentially non-dispositional

and non-structural features of subjective experience; (ii) if fundamental categorical properties are like the properties we find in subjective experience, then they are nonphysical; or (iii) if fundamental categorical properties are an “unknown x” (to borrow a criticism of Kant’s noumena from Nietzsche 1882, § 54) that cannot be defined or properly understood, then they are of no explanatory use whatsoever. None of these routes helps answer the panpsychic challenge, and it is hard to see what other route could be available for a Russellian physicalist.

Consider the first proposal, that the Russellian physicalist’s fundamental categorical properties are closely related to some non-categorical property or other. Locke’s categorical solidity—extensively discussed by Pereboom (2011)—is such a candidate. Categorical solidity is the fundamental categorical ground for dispositional solidity—Locke was thinking of simple material objects as solid corpuscles of some sort. Even assuming we have a conceptual grasp on what it means to be categorically solid, positing such a property seems not to help solve the hard problem. This is because positing fundamental categorical solidity still allows the conceivability argument to arrive at its conclusion unhampered: we seem capable of conceiving of a world which duplicates the categorical solidity of our world, yet which has different or no phenomenal properties. Categorical solidity—understood as a property especially suited to generate non-categorical solidity—offers no explanation for why we are conscious.

Primitive color (this could really be any secondary quality, like sound or smell), on the other hand, seems to help a bit with explaining consciousness. Visual experience seems to be constituted by sensations of color. Perhaps if the fundamental physical world were essentially and intrinsically colorful, this would help explain visual experience.

Of course, primitive color looks an awful lot like a fundamentally mental property, if this means primitive *phenomenal* color. Such a view is unacceptable for a Russellian physicalist: if the categorical role-filler for a quark (or whatever) is a little bit of color experience, then physicalism is false. Yet, if primitive color is meant in some other sense, like the categorical ground for a surface reflectance property or the categorical ground for a wavelength property, then this proposal runs into the same issue as that faced by categorical solidity: categorical properties that are defined relative to their non-categorical cousins seem to be of no use in solving the hard problem. Primitive color additionally has the issue that it seems to only help explain visual experience. How could a collection of primitive color properties yield human-level auditory or olfactory experiences? A complete mystery, in my eyes.¹¹

Finally, the third proposal says that fundamental categoricals are like Kant’s noumena, an “unknown x”. The Aristotelians’ “prime materiality”—if it means

¹¹ Of course, other primitive sensory or sensory-like properties can be posited, such that there is primitive color, and primitive sound, and primitive smell, and so on for all the sensory modalities. In my eyes, this further illustrates how ridiculous the primitive color proposal is: there would have to be as many types of fundamental categorical properties as there are sense modalities (and likely more—one for every kind of color, sound, smell, etc.), and somehow they would have to be employed in the brain in a way that selectively instantiates appropriate sensations in appropriate physical contexts. I find such a view both ugly and implausible.

anything other than just “materiality” or “physicality” (Aristotle apparently meant “pure potentiality”, which is unclear to me)—might be in this category, since it is vague what sort of property this term is supposed to refer to. This view says that we have an understanding of non-fundamental phenomenal properties that depend on fundamental categoricals, but this gives no insight on the nature of the categorical dependence base. Pereboom prefers something like this route—though he additionally says that fundamental categoricals are protophenomenal, which I reject—and of the three options on offer so far, it seems the best.

Yet this proposal must ultimately be discarded as well: the panpsychic challenge asserts that we can conceive of no physical categorical properties which can explain subjective experience. The “unknown x” response admits this. This response likely requires rejecting the first premise of the argument, which says that to solve the hard problem, the posited categoricals must explain human subjective experience. Though this might be a viable way to go, to me the first premise looks strong. If it is true that of all the properties we can conceive, only mental or protomental properties help explain consciousness, then this constitutes a strong reason to favor panpsychism or panprotopsychism over Russellian physicalism.

Luckily, there is another way. We might be able to describe some features of fundamental categorical properties that would help explain consciousness—and which are not themselves mental or protomental—without committing to a more specific characterization of fundamental categoricals. This would avoid the trilemma presented: these features of fundamental categoricals need not be especially tied to non-categorical properties, nor to mentality, and these features would give our concepts of fundamental categorical properties intelligible content. What features of categoricals might do this work?

It will help to list some various important features of phenomenology.

Efficacious phenomenal properties are causally efficacious

Intrinsic phenomenal properties are absolutely intrinsic (i.e. do not entirely depend on more fundamental properties that are dispositional or structural)

Ineffable understanding of a phenomenal property cannot be communicated to a person who has not instantiated that property

Subjective phenomenal properties are properties of mental subjects

Phenomenal phenomenal properties essentially have a felt character (this is nearly a tautology).

I will take it for granted that these are all features of phenomenal properties—after all, Russellian physicalists are (to use the terminology of Frankish 2016) strong realists about consciousness, and want to show how all of the typical dualist intuitions can be satisfied without sacrificing physicalism. A Russellian physicalist is not a moderate illusionist, who says that our typical conception of a phenomenal property is misguided in some deep way. Since dualists typically attribute all of these features to subjective experience, the strongest form of Russellian physicalism ought to as well.

Efficacy of phenomenal properties can be obtained if phenomenal properties have dispositional features, and those dispositional features are intimately linked

to their categorical (i.e. phenomenal) features. If phenomenal properties have no causal dispositions, then they are epiphenomenal—this is no good.¹² If dispositional and categorical features of phenomenal properties are only loosely related, then the painfulness of pain (which is a categorical feature) would only contingently cause avoidance behavior (a dispositional feature). The stance described here is opposed to the version of Russellian monism that Alter and Coleman (forthcoming) call “compatibilism”, which says that categorical properties might be causally efficacious even if there is not a tight modal connection between the categorical and dispositional. In my eyes, this view becomes especially unattractive when applied to human phenomenology, and not just fundamental metaphysics—such a view must say that our world’s pain can accompany the disposition to attraction behavior.¹³ Yet it seems absurd that the very feeling of pain that we experience—which is *bad*—could consistently dispose a person toward attraction behavior. Somehow the painfulness of pain must necessarily be related to avoidance behavior—despite the commitment or Russellian physicalists to worlds where *avoidance behavior and painfulness come apart!*

This can be accomplished if the dependence base for phenomenal properties necessitate dispositions: every world which duplicates our categorical properties also necessarily duplicates our dispositions. This allows phenomenal pain, which depends on fundamental categorical properties, to cause avoidance behavior *in virtue of its categorical nature*. Importantly, the necessitation relationship between categoricals and dispositions must only be one-directional if Russellian physicalism is to maintain its answer to the conceivability argument: categoricals necessitate dispositions, and not vice versa. This is because, if dispositions were to necessitate categorical properties, then every world which instantiates a disposition to avoid bodily harm would necessarily be one with the same phenomenal pain properties as our world. Yet, as I explained in the previous section, Russellian physicalists require the conceivability of our world’s dispositional properties being instantiated without our world’s phenomenal properties (and they accept the move from conceivability to possibility).

The view on offer here can be better understood via analogy: imagine a mouse-trap. Suppose that the properties of *being cheese* and *being peanut butter* are categorical, and the property of *being bait* is dispositional. Everyone should agree that both the properties of *being cheese* and *being peanut butter*, along with the fact that cheese and peanut butter are desirable for mice, necessitate *being bait*, but *being bait*

¹² My goal at this point in the paper is to show that Russellian physicalism is consistent with some basic requirements for mental causation. Toward the end of the paper, I will sketch out some more specific Russellian physicalist accounts of causation.

¹³ I use pain as an example of a phenomenal property which has an obvious behavioral requirement. Some may find my selection problematic on the grounds that some phenomenal properties do not seem to have a specific behavioral requirement—e.g. seeing the red of an apple seems not to necessitate any particular behavior. I nonetheless feel that at least some phenomenal properties should necessitate some behavioral dispositions, which is why I focus on pain in this paper. It also seems reasonable that *clusters* of phenomenal properties should necessitate behavioral dispositions, such as the phenomenology of seeing an apple eliciting the disposition to make the appropriate hand behavior to grasp the apple (if the apple is desired).

does not specifically necessitate *being cheese* or *being peanut butter*. This concrete example shows how a categorical property might determine a dispositional property, and not vice versa. Further, it serves as an analogue for the conceivability argument: assume that an actual world mousetrap uses peanut butter for its bait. We can conceive of a counterfactual mousetrap realizing its bait properties without realizing its peanut butter properties, just as we can conceive of the dispositions associated with phenomenology, yet no or different phenomenology.¹⁴

Intrinsicness of phenomenal properties is trivial to generate from fundamental categoricals: fundamental categorical properties are absolutely intrinsic, and anything constituted by them will be absolutely intrinsic too.¹⁵ Note that phenomenal properties can depend on both categorical and non-categorical properties while still being absolutely intrinsic, and so might depend both on fundamental categoricals and e.g. their dynamic arrangement in spacetime or causal relationships.

The ineffability of phenomenology is a common trope of dualism. I'll take this to mean that understanding of a phenomenal property cannot be properly communicated to someone who has not herself instantiated that property. This is supposed to capture Block's (1978, p. 281) claim about consciousness—stealing a line from the musician Louis Armstrong—that “if you got to ask, you ain't never gonna get to know”. In contrast, a property like *being a planet* is effable in this sense, since it seems that it can be fully understood by someone who does not personally instantiate that property. The ineffability of phenomenal properties goes hand-in-hand with their subjectivity (the feature of phenomenal properties that they are properties of mental subjects). This is because the ineffability of a phenomenal property piggybacks on the unique epistemic access that a subject of experience has to that property.

Phenomenal properties can be uniquely accessible to mental subjects if direct epistemic access to *any* categorical property is closed to any being who is not the

¹⁴ Some Russellian monists hold that there are possible worlds which have not just different categorical properties from our world, but no categorical properties at all (see Morris 2016). They also hold that such worlds could be alike in all non-categorical respects to our world. This is disanalogous to the mousetrap metaphor, since it seems inconceivable that a part of a mousetrap could instantiate the property of *being bait* without some lower-order role-filler. However, in my eyes, Russellian monists lose nothing by giving up the view that non-categorical-only worlds are possible, since the Russellian answer to the conceivability argument can be secured without this commitment. Further, denying the possibility of such worlds strengthens the motivation for Russellian monism, since it would be metaphysically impossible for a world to only instantiate the structural and dispositional properties described by science without there also being categorical role-fillers.

¹⁵ If absolute intrinsicness is only inherited by non-fundamental properties which are *constituted* by fundamental categorical properties, then non-fundamental properties which depend on fundamental categorical properties without being constituted by them cannot have absolute intrinsicness. This seems to rule out functional role properties from being identical to phenomenal properties under Russellian physicalism, since functional role properties are generally recognized to depend on their lower-order realizers without being constituted by them. A Russellian physicalist might still endorse realizer functionalism, and say that the actual-world realizers of appropriate functional role descriptions are identical to phenomenal properties. What exactly ‘constitution’ means remains open—my hope is that this openness allows Russellian physicalism to be consistent with a range of views on the nature of non-fundamental categorical properties. See the discussion at the end of the next section for some options.

possessor of that property. This view says that the categorical role-filler of an electron can only be known about (*per impossibile*) by the electron. On the assumption that phenomenal properties are categorical, a phenomenal property can only be known by the very person who possesses the phenomenal property. This epistemic feature of categoricals is not mentality-specific—since *any* categorical property (mental or otherwise) can only be known about by the entity (mental or otherwise) which possesses that property—and so avoids making fundamental categoricals mental or protomental in a physically problematic sense. The ineffability of phenomenal properties is thus ensured, since understanding of a phenomenal property could not be properly communicated to someone who is epistemically closed to that property.¹⁶

It would be helpful to compare epistemic access to categorical properties with epistemic access to non-categorical properties. Abstracta aside, it seems that we only acquire knowledge of non-categorical properties via their effects on our sensory systems or on measuring devices that extend the capacities of our sensory systems. We know about the hardness of a desk—a dispositional property—in virtue of the perceptible effects of that hardness. By definition, categorical properties are not exhausted by their dispositional features (if any). The effects of a categorical property on our perceptual system is insufficient to tell us everything about that categorical property, since such effects suffice only to inform us about dispositions. We know about the categorical features of our brains because our conscious minds are identical to those categorical features, and we cannot know about the categorical features of desks or electrons because our minds are not desks or electrons.

One might further like to know how the leap is made from fundamental non-subjects to the human level of subjects—this is a version of the combination problem (see Goff 2006; Coleman 2013; Chalmers 2016, Montero 2016). On this I defer to McClelland (2013), who builds on the work of Kriegel (2009): to be a subject, one must self-represent. This does not require higher-order representations (which McClelland and Kriegel explicitly oppose as the basis for explaining subjectivity): a phenomenal state might be a self-representing state, which does not require an additional higher-order state to represent it. On the widely-held assumption that representation depends on non-categorical properties (representation, like cognition more generally, is a scientifically scrutable phenomenon), it follows that there is no deep worry about generating subjects from fundamental categorical properties. This is because all one needs for subjectivity are scientifically scrutable structural or dispositional properties.

Finally, how do non-phenomenal fundamental categoricals generate higher-level properties that have a felt character (the phenomenal-ness of phenomenal

¹⁶ One might wonder whether we can acquire knowledge of the phenomenal properties of non-humans through instantiation of qualitatively identical phenomenal properties, e.g. by somehow generating bat-like experiences in a human. I am unsure whether this should count as knowledge, and perhaps it depends on whether those bat experiences were generated through chance (as in a dream), or through a more deliberate and reliable process, like neurosurgery. Even if knowledge of bat phenomenology can be acquired through neurosurgery (or something similar), the ineffability of phenomenology is still secure, since a person still only knows about the phenomenology of a bat by having the experiences of a bat.

properties)? My answer builds on the explanations given for ineffability and subjectivity: non-fundamental categoricals are type-indexed to the sort of being they are properties of. The categorical properties of a fried egg are *fried-egg-like*, whereas the categorical properties of a subject are *subject-like*—and only the latter has a felt character. The different types of non-fundamental categoricals depend on both their constituent categorical and non-categorical properties. Unfortunately, this account does not explain exactly how fundamental categoricals give rise to non-fundamental phenomenology. Since a Russellian physicalist admits that alternative fundamental categoricals in different possible worlds would not generate phenomenology, one would like to know what exactly it is about our world's categoricals that makes them phenomenology-apt.

While I do not have an answer, I think that Russellian physicalism is no worse off in this regard than panpsychism or (even more so) panprotopsyism. This is because both panpsychists and panprotopsyists must stipulate that the properties they posit are apt, given appropriate conditions, for generation of human-level phenomenology. Panpsychists posit bits of mind dust *with totally non-human phenomenology* that yield human consciousness, and panprotopsyists likewise posit bits of proto-mind-dust with a special aptness to produce human consciousness. In both cases, it is stipulated that the fundamental mental or protomental properties are apt to generate human phenomenology under appropriate conditions, and this is the end of the explanation. Sometimes the specific aptness conditions for mentality are described, as when Mørch (2018b) discusses the relationship between panpsychism and Tononi's (2007) integrated information theory, but a Russellian physicalist can endorse the same conditions. All that distinguishes Russellian physicalism from e.g. panprotopsyism is that Russellian physicalism says that fundamental categorical properties are capable of generating both phenomenal and non-phenomenal higher-level categorical properties, and panprotopsyism says that fundamental categoricals suffice only to generate phenomenal high-level categoricals. Thus, Russellian physicalism is no less explanatorily adequate than a non-physicalist option like panprotopsyism.

The features of fundamental categoricals I have identified are a bit schematic, but in my eyes that is both to be expected and not significantly worrisome. Categoricals, though inscrutable, need not be thought of only as an "unknown x". A more robust conception of fundamental categoricals and the categoricals that depend on them allows us to answer the panpsychic challenge: there are non-mental properties that we can conceive of which may generate full-blooded human phenomenal properties, though the what-it-is-likeness of phenomenal properties remains quite mysterious under any version of Russellian monism.

3 Support for Super-Dispositional Russellian Physicalism

In section one I described and defended two approaches to Russellian monism: super-categorical Russellian monism and super-dispositional Russellian monism (henceforth I will only discuss the physicalist versions of these two views). In section two, I responded to the panpsychic challenge by describing some non-mental

features of categorical properties which help explain various features of phenomenal properties. Now, I will look at how these projects fit together.

Recall that super-categorical Russellian physicalism says that fundamental categorical properties are super-categorical, and that they can swap with each other but not with higher-level categorical properties. This is distinct from—but not inconsistent with—super-dispositional Russellian physicalism, which says that scientifically scrutable properties are super-dispositional. Is there a reason to prefer one of these routes over the other? Or should a Russellian endorse both views, saying that that science describes only super-dispositional properties *and* that fundamental categorical properties are super-categorical?

The relevant deciding factor is the feature of fundamental categorical properties that explains the efficacy of phenomenal properties. In the previous section, I argued that phenomenal properties are only causally efficacious if they have dispositional features, and their dispositional features are necessarily tied to their categorical features. For example, the feeling of pain should determine the disposition to engage in avoidance behavior, and this determination relationship should not be metaphysically contingent. This is secured if both phenomenal categorical properties and their dispositions are metaphysically necessitated by fundamental categorical properties. If so, then any world which duplicates our fundamental categorical properties will also duplicate the dispositions associated with phenomenal properties, as well as the what-it-is-likeness of those phenomenal properties.

There is a further question regarding the *nature* of the necessitation relationship between categorical properties and non-categorical properties. There are two options for Russellian physicalists: laws-based necessitation, or brute necessitation. If the former, then categorical properties metaphysically necessitate non-categorical properties in virtue of synchronic laws of nature, such that every world containing a particular categorical property must additionally contain a law to generate a non-categorical property. Such a law can necessarily accompany a categorical property if that property is partially defined relative to a law, so that some particular property P can only exist in worlds with some particular law L.¹⁷ Brute necessitation, on the other hand, does not require laws to connect categorical properties to necessitated non-categorical properties. This other sort of necessitation is primitive and ultimately inexplicable.

There are reasons to prefer one or the other sort of view, e.g. parsimony favoring brute necessitation, and perhaps something like a principle of sufficient reason favoring law-based necessitation, but I think Russellian physicalism is consistent with either sort of necessitation. So long as fundamental categorical properties necessitate appropriate non-fundamental categorical and dispositional properties in all worlds in which those fundamental categorical properties occur, the efficacy of phenomenology is secured. Both laws-based and brute necessitation satisfy

¹⁷ It may be hard to see why categorical properties should be partially defined relative to laws which necessitate non-categorical properties, since it seems that categorical properties themselves are distinct from laws and from necessitated properties. While I admit that this may be an issue for laws-based accounts of necessitation, it is not a special problem for Russellian physicalism.

this requirement. If either of these routes work, then epiphenomenalism can be avoided.

Given the aforementioned understanding of what is required for phenomenal properties to be causally efficacious—that fundamental categorical properties necessitate both phenomenal properties and their dispositions—super-categorical Russellian monism ends up disallowing that phenomenal properties are efficacious. This entailment is nearly trivial, given how super-categorical Russellian physicalism is defined and the conditions required for causal efficacy of the mental. Recall that fundamental super-categorical properties are metaphysically capable of swapping without affecting any higher-level dispositional properties, such as pain avoidance behavior. If swapping fundamental categoricals metaphysically necessitates a change to higher-level phenomenal properties, it follows that the relationship between e.g. the feeling of pain and avoidance behavior is unacceptably loose, since it would be metaphysically contingent that avoidance behavior accompanies pain.

Under the assumption that causal efficacy of phenomenal properties requires necessitation of phenomenology by fundamental categorical properties, are there any options for a super-categorical Russellian physicalist? A super-categorical Russellian physicalist might respond that only non-categorical properties determine which particular phenomenal property is instantiated. On this view, all that is required for the feeling of pain would be fundamental categorical properties *of any kind whatsoever*, as well as some specific neural or functional state which determines what particular phenomenal property is instantiated. It would follow that swapping categoricals would not affect the painful feeling of pain. Unfortunately, this response will not work for a Russellian physicalist: as I have already pointed out, Russellian physicalists are committed to the view that swapping fundamental categoricals entails a change to phenomenal properties. This is what allows a Russellian physicalist to answer the conceivability argument: when we conceive of zombie worlds, we are conceiving of worlds with different or no fundamental categorical properties. If specific phenomenal properties are not determined by fundamental categorical properties, then the motivation for Russellian physicalism is undercut—no good!

This means that super-categorical Russellian physicalism is left without a way to avoid epiphenomenalism. Super-dispositional Russellian physicalism, on the other hand, is consistent with a weak form of dispositional essentialism (Bird 2005, 2007) which super-categorical Russellian physicalism is inconsistent with. Weak dispositional essentialism says that properties have their dispositional features essentially, such that every world which duplicates a property of our world also duplicates whatever dispositional features it has in our world. Under this view, a Russellian physicalist can say that whatever dispositional features our world's categorical properties possess would also be instantiated in any world which duplicates those categorical properties.

Weak dispositional essentialism of this sort allows for efficacy of phenomenal properties, since the relationship between the categorical and dispositional features of phenomenal properties would be modally tight. It would be metaphysically impossible for e.g. avoidance behavior to not accompany the feeling of pain. This

is because the fundamental categorical base for phenomenal pain would necessitate both the painful feeling as well as the disposition to avoid pain-causing stimuli.

Importantly, weak dispositional essentialism is different from the strong sort, which says that the necessitation relationship is two-directional between dispositional and non-dispositional features of a property. The upshot of strong dispositional essentialism for Russellian physicalism is that categoricals would necessitate dispositions and dispositions would necessitate categoricals. As I pointed out in the previous section, such a strong form of dispositional essentialism renders the Russellian physicalist answer to the conceivability argument unavailable: under strong dispositional essentialism, we would not be able to properly conceive of worlds with different categorical properties than our world, yet the same non-categorical properties.

Strong dispositional essentialism is employed by Alter and Coleman (2019, forthcoming) to describe “necessitarian” Russellian monism. Alter and Coleman claim that necessitarian Russellian monism provides a counter to an argument from Howell (2015), which concludes that categorical properties are not causally efficacious. Howell’s argument says that if categoricals necessitate dispositions, and not vice versa, then there could be a world with two different categorical grounds for one and the same dispositional property. This seems to render the categorical ground for that property causally irrelevant. Alter and Coleman respond that necessitarian Russellian monism disallows this possibility, since any given disposition will necessarily be matched with a particular categorical ground. They further argue that necessitarian Russellian monism is not really so unappealing, even though it seems unable to answer several anti-physicalist arguments—like the conceivability or knowledge argument—in the way that non-necessitarian Russellian monism can.

I suspect that Alter and Coleman are wrong on this, and that necessitarian Russellian monism loses most of the appeal that is supposed to motivate the view. Further, it is not clear to me why it would be problematic for two different categoricals to ground one and the same dispositional property in some world. In my view, Russellian monists must countenance the possibility of different categorical grounds for a single dispositional property in *different* worlds, so why think there is a problem with this occurring in a single world? If this is right, then Alter and Coleman’s response to Howell’s argument is made irrelevant. Regardless, even if Alter and Coleman are right, *necessitarianism is consistent with super-dispositional Russellian physicalism*. This is because, as already stated, super-Russellian physicalism does not carry a specific commitment to the necessitation relationship between categorical and non-categorical properties, and can be wedded to either weak or strong dispositional essentialism.

However, super-categorical Russellian physicalism is inconsistent with both strong and weak dispositional essentialism. Given that epiphenomenalism is an unattractive view, and that weak or strong dispositional essentialism is required to secure the causal efficacy of the mental, it follows that super-dispositional Russellian physicalism is preferable to super-categorical Russellian physicalism. Though there is no logical barrier that prevents a Russellian physicalist from endorsing both super-categorical and super-dispositional options, she should not, lest embrace epiphenomenalism.

Unfortunately, a problem arises for Russellian physicalism which is similar in many respects to the well-known causal exclusion argument against non-reductive physicalism (Kim 1992).¹⁸ Russellian physicalists say that properly related fundamental categorical properties necessitate both non-fundamental categorical and non-categorical properties. Under the assumption that properties like *being solid* are non-categorical, and do not depend on anything non-categorical until we get down to the fundamental level—since solidity seems to entirely depend upon the dispositional property of having tight bonds among its molecular constituents, which in turn depends upon the dispositional properties of electrons and other atomic constituents—it seems that there is no room for non-fundamental categorical properties to do any causal work. However, as I have described the view, Russellian physicalism additionally requires that non-fundamental categorical properties are causes of effects, such that e.g. phenomenal pain causes human-level avoidance behavior. One can seemingly tell the whole causal story without mentioning non-fundamental categorical properties, just as traditional reductionists like Kim seem to be able to tell the whole story about human behavior without mentioning functionally-characterized mental properties. Non-fundamental categorical properties appear to be causal danglers.

This problem has several possible answers which are analogues to the standard answers of the more familiar exclusion problem against non-reductive physicalism, and any of them can *prima facie* be endorsed by a Russellian physicalist.¹⁹ One can say that (i) there is non-problematic overdetermination of behavior by supervening non-fundamental categoricals, similar to the compatibilism of Bennett (2003, 2008); (ii) that non-fundamental causation only depends on non-fundamental categorical properties, similar to the proportional causation of Woodward (2003); or that non-fundamental categoricals are identical to conjunctions of fundamental categoricals and their relations, as per reductionism. I think the compatibilist route is fairly straightforward—this view says non-fundamental categoricals overdetermine their effects along with sufficient non-categorical causes in a non-problematic way, in virtue of appropriate necessitation relationships between categoricals and non-categoricals—so let me describe the other two routes in a bit more detail.

A Russellian physicalist who believes in proportional causation would say that causes are proportional to their effects, and that fundamental categorical properties are not proportional to e.g. behavioral effects. One might wonder why non-categorical properties of e.g. neurons are not the causes of behavior, rather than non-fundamental categorical properties of neurons. A proportional causation Russellian monist might urge that non-categorical properties are not themselves causes, but only descriptions of the causal powers of categorical properties. If so, then non-categorical properties are themselves epiphenomenal. There may be other ways to articulate

¹⁸ Thanks to an anonymous Erkenntnis referee for pushing me on this.

¹⁹ Upon investigation, it may turn out that one or more of these solutions to the Russellian-monism-specific exclusion problem are barred to a Russellian physicalist. Unfortunately, such an investigation is beyond the scope of the present paper, and it suffices for present purposes that each of these solutions is at least *prima facie* available to a Russellian physicalist.

this sort of view (I admit that it sounds strange to say that dispositions are epiphenomenal), but I believe that such a project is outside my present scope.

A reductionist would say that what it means to be a non-fundamental categorical property is to be identical to some conjunction of fundamental properties and their relations. Fundamental categorical properties individually necessitate low-level dispositions, e.g. subatomic dispositions, and collectively necessitate higher-level dispositions, e.g. pain avoidance dispositions. When telling the story about ascending to higher levels of dispositions, there is an implicit assumption for the reductionist that this is simultaneously a story about more inclusive sets of fundamental categorical properties and their relations, e.g. the constellation of fundamental categorical properties which underlie neural properties. On this view, categorical properties of quarks necessitate e.g. spin and color, and categorical properties of atoms necessitate e.g. dispositions to form covalent bonds. A reductionist can say that spin and color of quarks are fine-grained non-categorical properties necessitated by individual categorical properties, and that atomic dispositions are more coarse-grained non-categorical properties necessitated by a more inclusive set of related fundamental categorical properties. Under the assumption that there is no problem with the causal power of a fundamental categorical property, a reductionist can claim that there is similarly no problem for non-fundamental categorical properties, which are (again) identical to conjunctions of fundamental categorical properties and their relations.²⁰

4 Conclusion

I have argued that the criticisms levied against Russellian monism by Hiddleston fail, but that they illuminate important differences between super-categorical and super-dispositional Russellian monism. This was followed by an assessment of the panpsychic challenge against Russellian physicalism—that only fundamental phenomenal or protophenomenal properties are capable of explaining human phenomenal properties—where I argued that various mentality-neutral features of categoricals can be described which aid in understanding how categorical properties give rise to higher-level phenomenal properties. Finally, I showed that the feature of fundamental categoricals which explains the efficaciousness of phenomenal properties provides a strong reason to prefer super-dispositional Russellian physicalism over super-categorical Russellian physicalism. Hopefully by addressing Hiddleston's arguments and the panpsychic challenge, and using these responses to flesh out Russellian physicalism, I have filled in some of the missing details to an otherwise attractive position.

²⁰ One might object that the reductionist route *eliminates* non-fundamental categorical properties, since there seems to be nothing more in this ontology than fundamental categoricals and their relations. However, I think reduction of this sort is not elimination: a non-fundamental categorical property N might be identical to the set $(F1, F2, R_{F1,F2})$, which is the conjunction of two fundamental properties and a relation between them. This set is not identical to any individual fundamental property—so non-fundamental categorical properties are not eliminated under this view, if one is a realist about these sets.

References

- Alter, T., & Coleman, S. (forthcoming). *Russellian monism and mental causation*. Noûs.
- Alter, T., & Coleman, S. (2019). Panpsychism and Russellian monism. In W. Seager (Ed.), *The Routledge companion to panpsychism*. New York: Routledge.
- Alter, T. (2016). The structure and dynamics argument against materialism. *Noûs*, 50(4), 794–815.
- Alter, T., & Nagasawa, Y. (2012). What is Russellian monism? *Journal of Consciousness Studies*, 19(9–10), 67–95.
- Armstrong, D. (1961). *Perception and the physical world*. London: Humanities Press.
- Bennett, K. (2003). Why the exclusion problem seems intractable and how, just maybe, to tract it. *Noûs*, 37(3), 471–497.
- Bennett, K. (2008). Exclusion again. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced: New essays on reduction, explanation, and causation* (pp. 280–307). Oxford: Oxford University Press.
- Bird, A. (2005). Laws and essences. *Ratio*, 18, 437–461.
- Bird, A. (2007). *Nature's metaphysics: Laws and properties*. Oxford: Oxford University Press.
- Block, N. (1978). Troubles with functionalism. *Minnesota Studies in the Philosophy of Science*, 9, 261–325.
- Brown, C. D. (2017). A properly physical Russellian physicalism. *Journal of Consciousness Studies*, 24(11–12), 31–50.
- Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.
- Chalmers, D. J. (2003). Consciousness and its place in nature. In S. P. Stich & T. A. Warfield (Eds.), *Blackwell guide to the philosophy of mind* (pp. 102–142). Oxford: Blackwell.
- Chalmers, D. J. (2010). *The character of consciousness*. Oxford: Oxford University Press.
- Chalmers, D. J. (2015). Panpsychism and panprotopsychism. In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the physical world: Perspectives on Russellian monism* (pp. 246–276). Oxford: Oxford University Press.
- Chalmers, D. J. (2016). The combination problem for panpsychism. In G. Brüntrup & L. Jaskolla (Eds.), *panpsychism* (pp. 179–214). Oxford: Oxford University Press.
- Coleman, S. (2013). The real combination problem: Panpsychism, micro-subjects, and emergence. *Erkenntnis*, 1, 1–26.
- Coleman, S. (2015). Neuro-cosmology. In P. Coates & S. Coleman (Eds.), *Phenomenal qualities: Sense, perception, and consciousness* (pp. 66–102). Oxford: Oxford University Press.
- Coleman, S. (2016). Panpsychism and neutral monism: how to make up one's mind. In G. Brüntrup & L. Jaskolla (Eds.), *Panpsychism: Contemporary perspectives* (pp. 249–282). New York: Oxford University Press.
- Dennett, D. (1991). *Consciousness Explained*. Penguin Books.
- Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11–12), 11–39.
- Goff, P. (2006). Experiences don't sum. *Journal of Consciousness Studies*, 13(10–11), 53–61.
- Goff, P. (2009). Why panpsychism doesn't help us explain consciousness. *Dialectica*, 63(3), 289–311.
- Heil, J. (2005). Dispositions. *Synthese*, 144(3), 343–356.
- Hellman, G. (1985). Determination and logical truth. *Journal of Philosophy*, 82, 607–616.
- Hellman, G., & Thompson, F. (1975). Physicalism: Ontology, determination and reduction. *Journal of Philosophy*, 72, 551–564.
- Hiddleston, E. (2019). Dispositional and categorical properties, and Russellian Monism. *Philosophical Studies*, 176(1), 65–92.
- Howell, R. J. (2009). Emergentism and supervenience physicalism. *Australasian Journal of Philosophy*, 87(1), 83–98.
- Howell, R. J. (2015). The Russellian monist's problems with mental causation. *Philosophical Quarterly*, 65, 22–39.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127–136.
- Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy*, 83(5), 291–295.
- Kim, J. (1992). The nonreductivist's trouble with mental causation. In J. Heil, & A. R. Mele (Eds.), (pp. 189–210). Oxford: Oxford University Press.

- Kirk, R. (1974). Zombies vs materialists. *Proceedings of the Aristotelian Society*, 48, 135–152.
- Kriegel, U. (2009). *Subjective consciousness: A self-representational theory*. Oxford: Oxford University Press UK.
- Kripke, S. (1980). *Naming and necessity*. Cambridge: Harvard University Press.
- Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64, 354–361.
- Levine, J. (2001). *Purple Haze*. Oxford: Oxford University Press.
- Lewis, D. (2009). Ramseyan humility. In D. B. Mitchell & R. Nola (Eds.), *Conceptual analysis and philosophical naturalism* (pp. 203–222). Cambridge: MIT Press.
- Martin, C. B. (1997). On the need for properties: The road to pythagoreanism and back. *Synthese*, 112(2), 193–231.
- McClelland, T. (2013). The neo-Russellian ignorance hypothesis: A hybrid account of phenomenal consciousness. *Journal of Consciousness Studies*, 20(3–4), 125–151.
- Melnyk, A. (1997). How to keep the “physical” in physicalism. *Journal of Philosophy*, 94(12), 622–637.
- Melnyk, A. (2002). Physicalism. In S. P. Stich & T. A. Warfield (Eds.), *Philosophy and phenomenological research* (pp. 573–587). Oxford: Blackwell.
- Montero, B. (2005). What is the physical? In A. Beckermann (Ed.), *The Oxford handbook of philosophy of mind* (pp. 173–188). Oxford: Oxford University Press.
- Montero, B. (2010). A Russellian response to the structural argument against physicalism. *Journal of Consciousness Studies*, 17(3–4), 3–4.
- Montero, B. (2015). Russellian physicalism. In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the physical world: Perspectives on Russellian monism* (pp. 209–223). Oxford: Oxford University Press.
- Montero, B. (2016). What combination problem? In G. Brüntrup & L. Jaskolla (Eds.), *Panpsychism, Contemporary Perspectives* (pp. 227–43). Oxford: Oxford University Press.
- Montero, B., & Papineau, D. (2005). A defense of the via negativa argument for physicalism. *Analysis*, 65, 233–237.
- Mørch, H. H. (2018a). The evolutionary argument for phenomenal powers. *Philosophical Perspectives*, 31(1), 293–316.
- Mørch, H. H. (2018b). Is the Integrated Information Theory of Consciousness Compatible with Russellian Panpsychism? *Erkenntnis*, 84, 1065–1085.
- Morris, K. (2016). Russellian physicalism, bare structure, and swapped inscrutables. *Journal of Consciousness Studies*, 23(9–10), 180–198.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435–450.
- Papineau, D. (2001). The rise of physicalism. In C. Gillett & B. Loewer (Eds.), *Physicalism and its discontents* (pp. 3–36). Cambridge: Cambridge University Press.
- Papineau, D., & Spurrett, D. (1999). A note on the completeness of “physics.” *Analysis*, 59(1), 25–29.
- Pereboom, D. (2011). *Consciousness and the prospects of physicalism*. Oxford: Oxford University Press.
- Pereboom, D. (2013). Russellian monism and absolutely intrinsic properties. In U. Kriegel (Ed.), *Current controversies in philosophy of mind* (pp. 40–69). London: Routledge.
- Poland, J. (1994). *Physicalism, the philosophical foundations*. Oxford: Oxford University Press.
- Seager, W. (2006). The ‘intrinsic nature’ argument for panpsychism. *Journal of Consciousness Studies*, 13(10–11), 129–145.
- Smart, J. J. C. (1978). The content of physicalism. *Philosophical Quarterly*, 28, 339–341.
- Smith, A. D. (1993). Non-reductive physicalism? In H. M. Robinson (Ed.), *Objections to Physicalism* (pp. 225–250). Oxford: Oxford University Press.
- Stoljar, D. (2001). Two conceptions of the physical. *Philosophy and Phenomenological Research*, 62(2), 253–281.
- Strawson, G. (2006). Realistic monism—why physicalism entails panpsychism. *Journal of Consciousness Studies*, 13(10–11), 3–31.
- Tononi, G. (2007). The information integration theory of consciousness. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 287–299). Oxford: Blackwell.
- Wilson, J. (2006). On characterizing the physical. *Philosophical Studies*, 131(1), 61–99.
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
- Worley, S. (2006). Physicalism and the via negativa. *Philosophical Studies*, 131(1), 101–126.