

Against the Middle Ground: Why Russellian Monism is Unstable

1. Introduction

In recent years, many philosophers have embraced *Russellian monism* as an alternative to traditional physicalist and dualist views of consciousness (Maxwell 1979, Stoljar 2001, Strawson 2006, Montero 2010, Coleman 2015, Chalmers 2015, Goff 2017). Russellian monism represents a middle ground between physicalism and dualism, a refuge for those who reject physicalism for the usual reasons but wish to avoid the extravagancies of dualism. I argue that this middle-ground position is unstable. Those who reject physicalism on the basis of the usual epistemic arguments, such as the conceivability argument, should reject Russellian monism on similar grounds and embrace some form of dualism. Conversely, those unwilling to go all the way to dualism must reject the epistemic arguments against physicalism. In either case, these arguments cannot be used to motivate Russellian monism. For simplicity, I shall focus on the conceivability argument as a representative epistemic argument against physicalism, though everything I say could be adapted to other epistemic arguments against physicalism, such as Jackson's (1982) knowledge argument.

The plan for this paper is as follows: In §2, I briefly summarize the conceivability argument and make an observation that will prove relevant later on. The observation, very roughly, is this: because we are ignorant of the specific physical underpinnings of consciousness, the success of the conceivability argument depends on the plausibility of certain *general* epistemic gap principles of roughly the form “no truth of type *X a priori* entails any experiential truth,” such that we can know in advance that the truths about the physical underpinnings of consciousness will be truths of type *X*. In §3, I clarify the thesis of Russellian monism and distinguish various forms of the view. In §4, I argue that there are general epistemic gaps that

can be used to generate conceivability arguments against all forms of Russellian monism, gaps that are no less plausible than those implicitly invoked in the conceivability argument against physicalism. Finally, in §5, I argue that the link between conceivability and possibility assumed by the original conceivability argument makes Russellian monism vulnerable to a different conceivability argument, one involving “ghosts” rather than zombies. The operative principles of the conceivability argument therefore turn out to be no more friendly to Russellian monism than to physicalism.

2. The Conceivability Argument

The conceivability argument against physicalism begins with the premise that there are conceivable scenarios that perfectly resemble the actual world in all physical respects but differ from the actual world in experiential respects. The second premise of the argument says that if scenarios of this kind are coherently conceivable, then they are possible. It follows from these premises that the physical truths do not metaphysically necessitate the experiential truths. Since the truth of physicalism plausibly requires at least the supervenience of experiential truths on physical truths, we may conclude that physicalism is false. Following Chalmers (2002), we can formalize the argument as follows, letting “*P*” abbreviate the conjunction of all physical truths and letting “*Q*” abbreviate some experiential truth (e.g. “someone is conscious”):

- C1. $P \& \sim Q$ is conceivable.
- C2. If $P \& \sim Q$ is conceivable, then $P \& \sim Q$ is possible.
- C3. If $P \& \sim Q$ is possible, then physicalism is false.
- C4. Therefore, physicalism is false.¹

¹ Elsewhere Chalmers (1996, 2009) gives a more sophisticated version of the argument that makes use of the framework of epistemic two-dimensionalism. The two-dimensionalist conceivability argument is discussed in §5, but the more straightforward argument given here will suffice for current purposes.

The relevant notion of “conceivability” here is closely tied to apriority. We can say that a hypothesis is conceivable just in case it is a rationally coherent hypothesis, one that cannot be ruled out *a priori*, even given ideal rational reflection (Chalmers 2002). Let us say that there is an *epistemic gap* between *A* and *B* just in case *B* is not *a priori* entailed by *A*—that is, just in case the material conditional $A \supset B$ is not knowable *a priori*, even upon ideal rational reflection. Premise C1 therefore amounts to the claim that there is an epistemic gap between *P* and *Q*.

There are many objections that could be raised to this argument, most of which I shall not consider here. However, I do want to briefly mention one important objection to premise C1, since the standard response to this objection will be important later on. The objection, in brief, is that we are not physically omniscient—that is, we do not know what all is included in *P*—so we are in no position to know whether *P a priori* entails *Q* (Stoljar 2006). Even if we restrict our attention to physical truths about human brains, or, more narrowly, to the neuro-functional correlates of consciousness, there is much we do not know. Perhaps the facts about consciousness are not *a priori* entailed by any physical truths *we know*, or even by any (consistent) physical hypothesis we’ve yet entertained. But there is a great deal we don’t know, and a great many physical hypotheses we haven’t entertained. So what justifies our confidence that the physical truths, whatever they may turn out to be, will not *a priori* entail the truths about consciousness? Let us call this the “Ignorance Objection.”

The standard response to the Ignorance Objection says that, while we do not know, and indeed haven’t even entertained, all physical truths related to consciousness, we at least know that the relevant physical truths will be truths *of a certain type*, and we know that no truths of that type *a priori* entail experiential truths. (An analogy: we can arguably know that purely geometric truths never *a priori* entail positive truths about color. In that case, one might be

justified in holding that the geometric truths about the cup on my desk do not *a priori* entail that the cup is yellow, even if one does not know in any detail what the cup is like in geometric respects [Cutter forthcoming].) The justification for C1 will therefore rely on a pair of background assumptions of the following form:

(A) All physical truths are truths of type *X*.

(B) Truths of type *X* don't *a priori* entail experiential truths.

Claims with the form of (B) assert the existence of what I shall call a *general* epistemic gap, a gap between experiential truths and all truths of a certain general *type*. Together with a corresponding claim of form (A), this general epistemic gap will entail the existence of the *specific* epistemic gap between *P* and *Q* asserted by C1.

One way to fill out this schema would be to take type-*X* truths to be *non-experiential truths*. In that case, we'd motivate C1 by appeal to a general non-phenomenal/phenomenal gap:

Non-phenomenal/phenomenal gap: no collection of non-experiential truths *a priori* entails any experiential truth (cf. Chalmers 2015: 260).

A more common approach is to choose a value of “*X*” that corresponds more closely to the subject matter of physics. It is commonly held that physics is in the business of mapping the spatiotemporal, causal, and mathematical *structure* of the world and the way this structure changes over time. For this reason, defenders of the conceivability argument commonly take type-*X* truths to be truths about “structure and dynamics” (Chalmers 2002, Pereboom 2011, Alter 2016). Roughly speaking, truths about structure and dynamics (hereafter, “structural truths”) are truths expressible in a purely structural language, an austere language that includes only spatiotemporal, causal/nomic, mathematical, and logical expressions (Chalmers 2012). In other

words, the strategy under consideration responds to the Ignorance Objection by invoking the following general epistemic gap:

Structural/phenomenal gap: No set of structural truths *a priori* entails any experiential truth.

This principle is supposed to have a great deal of intuitive plausibility, and I shall not question this assessment. For example, it's supposed to be intuitively obvious that an experiential truth like "someone is having a phenomenally reddish experience" could never be *a priori* deduced by truths expressible in an austere structural language. Given the structural/phenomenal gap together with the assumption that all physical truths are structural truths (an assumption we shall revisit in our discussion of Russellian monism in §3), the truth of C1 follows.

I emphasize the conceivability argument's reliance on a general epistemic gap of this kind because, as I shall argue in §4, there are other general epistemic gaps of more-or-less comparable intuitive plausibility that create serious difficulties for Russellian monism. First, though, we must clarify the Russellian monist view and see how it attempts to sidestep the conceivability argument.

3. Russellian Monism

The idea that the content of physics is limited to claims about structure turns out to be one of the guiding ideas of Russellian monism, a view about the relationship between consciousness and the physical inspired by ideas in Russell's *The Analysis of Matter* and other works. Russell argued that physics characterizes matter in terms of its relational structure, but does not reveal its intrinsic nature. Of course, physics tells us that the fundamental physical entities have certain properties, such as mass and charge, but these ascriptions plausibly just amount to more relational information—specifically, information about how the entities in

question *causally* relate to other things. After all, the cash value of the claim that something has mass is just that it resists acceleration, attracts other massive things, and so forth. But this is just information about how it is disposed to *affect* and *be affected by* other things. We might put the point by saying that physics tells us about the causal dispositions or second-order role properties of the fundamental physical entities, but does not reveal the intrinsic qualities that serve as the categorical bases or realizers for these causal dispositions/role properties.

Let's use the term "quiddity" for these unknown qualities of matter. That is, quiddities are defined as the intrinsic properties that serve as the categorical bases for the fundamental causal dispositions characterized by physics. It's not obvious that there must be quiddities, for it's not obvious that there must be anything "underwriting" the structure revealed by physics (Mumford 2006; Ladyman & Ross 2007). But for the purposes of this paper, I shall assume that quiddities exist and that quidditative truths are in some good sense beyond the reach of the physical sciences. (This assumption is to my opponent's advantage, so it's fair in the present dialectical context.)

It will be useful to distinguish two senses of the term "physical property." Following Chalmers's (2015) helpful terminology, we'll say that *narrowly physical properties* exclude quiddities, and include only the relational/structural properties in terms of which physics characterizes matter, such as spatiotemporal properties and the second-order role properties associated with mass and charge (together with any properties definable in terms of these via logical/mathematical operations), and we'll say that *broadly physical properties* include quiddities in addition to all narrowly physical properties. We'll say that narrowly physical *truths* are truths concerning the instantiation of narrowly physical properties, and likewise, *mutatis mutandis*, for broadly physical truths. From here, we can define Russellian monism as the view

that macro-experiential truths—that is, truths about the conscious experiences of macroscopic entities like human beings—are grounded in broadly physical truths, but not (wholly) grounded in narrowly physical truths.

For the purposes of this paper, it will be useful to understand the terms “physicalism” and “dualism” in a way that sets them cleanly in opposition to Russellian monism. Physicalism, as I shall understand it, holds that macro-experiential truths are grounded in narrowly physical truths. Dualism, by contrast, holds that macro-experiential truths are fundamental, in the sense that they are not grounded in, or constitutively dependent on, anything more basic. Russellian monism can then be seen as a middle ground between physicalism and dualism. The Russellian monist agrees with the dualist against the physicalist that narrowly physical truths aren’t sufficient to ground macro-experience, but agrees with the physicalist against the dualist that macro-experience is grounded in more basic, lower-level facts.

Russellian monism comes in *panpsychist* and *panprotopsychist* forms. According to the former, quiddities are experiential properties. We can call these “micro-experiential properties,” since they characterize the fundamental microphysical entities.² According to the latter, quiddities are “proto-experiential” properties. Experiential properties are, roughly speaking, ways of being conscious, or determinates under the determinable property *consciousness*. Proto-experiential properties, by contrast, are non-structural properties which are not themselves experiential, but can ground experiential properties when appropriately combined (Chalmers 2015).

² Note that not all forms of panpsychism are forms of Russellian monism. Panpsychism can be defined as the view that the fundamental microphysical entities have experiential properties. To get from the bare thesis of panpsychism to Russellian monism, we need two further claims: (i) Microexperiential properties are quiddities (that is, they occupy the basic causal roles characterized by physics). (ii) Macro-experience is grounded in micro-experience (perhaps together with structural properties).

Russellian monists argue that their view, unlike physicalism, is invulnerable to the conceivability argument. After all, when one tries to imagine a creature that resembles you in physical respects while differing from you in experiential respects, the physical features one imaginatively holds fixed are presumably your *narrowly* physical features. Since the physical sciences don't tell us what things are like in quidditative respects, one is not in a position to imagine something just like you in all *broadly* physical respects. We therefore have no reason to suppose that there is an epistemic gap between the *broadly* physical truths and the experiential truths (Stoljar 2001, Alter & Nagasawa 2012, Chalmers 2015). Or so the argument goes. In the following section, we'll see that there are grounds for questioning this conclusion.

4. Russellian Monism's Epistemic Gaps

In §2, we saw that the conceivability argument relies on a general epistemic gap—typically, the structural/phenomenal gap or something in the vicinity. In this section, I argue that there are other plausible epistemic gaps that can be used to generate conceivability arguments against Russellian monism. Here I focus on three such gaps. The first, which has been emphasized in recent work by Goff (2009) and Chalmers (2015), targets panpsychist Russellian monism. The second and third, which have received much less (if any) attention, target both panpsychist and panprotopsychist forms of Russellian monism.

Let's begin with panpsychist forms of Russellian monism. The most famous problem for panpsychist Russellian monism is the *combination problem*, roughly the problem of saying how micro-experience combines to yield macro-experience. The canonical statement of the problem comes from William James:

Take a hundred [feelings], shuffle them and pack them as close together as you can (whatever that may mean); still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean. There would be a hundred-and-first-feeling there, if, when a group or series of such feelings were set up,

a consciousness belonging to the group as such should emerge. And this 101st feeling would be a totally new fact; the 100 feelings might, by a curious physical law, be a signal for its creation, when they came together; but they would have no substantial identity with it, nor it with them, *and one could never deduce the one from the others*, nor (in any intelligible sense) say that they evolved it. (James 1890/1981: 160, italics mine)

An interesting but rarely noted part of James's statement of the combination problem is his suggestion (italicized above) that panpsychism is burdened with an *epistemic gap*: the experience of a composite system could never be *a priori* deduced from the experiences and arrangement of its parts. The combination problem, or one form of it, can therefore be seen as raising the same problem for panpsychism that the conceivability argument raises for physicalism. And just as the original conceivability argument relies on a general epistemic gap, we can read James as appealing to a general epistemic gap to the effect that the experiences of one set of subjects never *a priori* entails the existence of a distinct subject. A bit more precisely, we can formulate the principle as follows (where a "subject" is defined as anything with experiential properties):

Subject/subject gap: for any experiential properties Q_1, \dots, Q_n (and any structural relation R), if it is conceivable that

(A) there exist some subjects $x_1 \dots x_n$ which instantiate $Q_1 \dots Q_n$, respectively (and stand in R),

then it is conceivable that: (A) holds and there are no subjects distinct from each of $x_1 \dots x_n$. (cf. Goff 2009; Chalmers 2015: 267)

This principle is quite plausible. Indeed, in terms of intuitive plausibility, this principle seems to be on a par with the structural/phenomenal gap that motivates the first premise of the original conceivability argument. To appreciate the plausibility of the subject/subject gap, imagine first an ordinary zombie, an unconscious creature that perfectly resembles a human being in all narrowly physical respects. Now "color in" its microscopic parts with any experiences you like.

Intuitively, no matter how you color in the parts, it remains conceivable that the system as a whole lacks consciousness. The same point can be illustrated by considering the problem of other minds within a panpsychist setting. The traditional problem of other minds says that no amount of (narrowly) physical information about an individual x (e.g. another human, or a bat, or a robot) will allow us to determine with certainty that x is conscious. But if this is correct, then surely our uncertainty would not be eliminated upon receiving further information about the experiences of things distinct from x , such as x 's microscopic parts. For example, if we cannot tell whether an advanced AI is conscious by learning about its physical and functional properties, then surely it will not help us to be told about any experiences that may be enjoyed by its capacitors and transistors.

Of course, the panpsychist Russellian monist can make an appeal to ignorance: we don't know what the experiences at the microscopic level are like, so we are in no position to know what may be *a priori* deducible from them. But it's not clear why this appeal to ignorance has more dialectical force than the Ignorance Objection to the original conceivability argument. Recall that my target is a Russellian monist who has rejected physicalism on the basis of the conceivability argument. The success of my argument therefore does not require that the subject/subject gap can be known beyond all doubt. All that is required is that the subject/subject gap should be more-or-less on a par with the structural/phenomenal gap (or whatever other general gap is used to motivate the first premise of the original conceivability argument) with respect to intuitive plausibility. For the panpsychist Russellian monist who is motivated by the conceivability argument must accept the structural/phenomenal gap (or something similar) while rejecting the subject/subject gap. But if, as I contend, they are antecedently approximately equally plausible, then this sort of differential treatment would be unjustified.

The subject/subject gap immediately entails the conceivability of “panpsychist zombies,” creatures that lack consciousness but perfectly resemble us in structural and micro-experiential respects (cf. Goff 2009, Chalmers 2015). If we take conceivability as a guide to possibility, we will infer that panpsychist zombies are possible and conclude that panpsychist Russellian monism is false. Of course, the panpsychist Russellian monist could resist the inference from conceivability to possibility, just as *a posteriori* physicalists resist the corresponding move in the original conceivability argument. But a Russellian monist who rejects physicalism on the basis of the conceivability argument is clearly in no position to offer this response.³

The subject/subject gap only threatens panpsychist forms of Russellian monism, not panprotopsychist forms. However, panprotopsychist Russellian monism founders on other epistemic gaps. It is sometimes said that panprotopsychism can be dismissed for the same reason as physicalism, namely, that it attempts to ground experiential truths in non-experiential truths. The implicit assumption here is that the rejection of physicalism is motivated by the non-phenomenal/phenomenal gap. Perhaps this is the best way to motivate the first premise of the original conceivability argument, in which case the operative gap would immediately yield an analogous objection to panprotopsychist Russellian monism. But as we’ve seen, this premise can also be motivated by appeal to a logically weaker gap, such as the structural/phenomenal gap. Since proto-experiential properties are non-structural by definition, the structural/phenomenal gap is no threat to panprotopsychism. Although I think it’s reasonably plausible that there is a general non-phenomenal/phenomenal gap, I want to focus on two other epistemic gaps which

³ Although the intuition pumps offered above presuppose that my panpsychist opponent holds a “bottom-up” view, according to which human consciousness is grounded in the consciousness of microscopic things, the argument given here applies equally to “top-down” panpsychist views which ground human consciousness in the consciousness of the cosmos. The latter view is defended by Nagasawa & Wager (2016), and a closely related view is defended by Goff (2017: ch. 9). For top-down panpsychist views, the relevant instance of the subject/subject gap would have $n = 1$ and would take Q_1 to be the total experiential property of the cosmos.

have not been discussed much (if at all), and which enjoy, by my lights, even greater intuitive plausibility than the non-phenomenal/phenomenal gap.

To introduce the first gap, let's first imagine that we're given a purely structural description of a conscious system such as a human brain, a description telling us that the system has a certain number of component parts that are spatially and causally related in certain ways (e.g. neurons in such-and-such spatial configuration, which are excited and inhibited by one another in certain specific ways). Given the structural/phenomenal gap, we will not be able to deduce the experiential properties of the system from this description. Suppose we're then told about the experiential properties (if any) possessed by the component parts. Given the subject/subject gap, we *still* won't be able to deduce the experiential properties associated with the system as a whole. Crucially, though, there doesn't seem to be anything special about the fact that we added information about the *experiential* properties of the component parts. It seems that *no* information about the intrinsic character of the component parts would allow us to deduce the experiential properties of the system as a whole. Whether the component parts are "colored in" with experiential qualities or non-experiential qualities, we won't be in a position to work out the experiential properties of the system as a whole. These considerations suggest that there is a general gap between "low-level" information about a system and the experiential properties associated with the system as a whole. Here we can define "low-level" information as information specifying the intrinsic character of the individual parts and the structural relations that hold among the parts. The relevant gap can be formulated a bit more carefully as follows:

Low-level/high-level gap: for any intrinsic properties F_1, \dots, F_n , and any structural relation R , if it is conceivable that

(A) there are some objects $x_1 \dots x_n$ which instantiate $F_1 \dots F_n$, respectively, and stand in R ,

then it is conceivable that: (A) holds and nothing of which $x_1 \dots x_n$ are proper parts is conscious.

Low-level/high-level gaps don't hold in general for high-level properties. For example, the shape of a composite object is *a priori* derivable from low-level information about the intrinsic properties and organization of its parts. But our concepts for experiential properties seem to be inferentially insulated from low-level information in a way that concepts for other properties are not.

The low-level/high-level gap makes trouble for both panpsychist and panprotopsychist forms of Russellian monism. According to the latter, macro-experience is grounded in non-experiential quiddities (perhaps together with structural properties). But quiddities, whether experiential or non-experiential, are supposed to be intrinsic properties of fundamental physical entities. Thus, given the low-level/subject gap, any form of Russellian monism will face an epistemic gap.

Again, my opponent can make an appeal to ignorance. Although no low-level intrinsic properties *that we know of* are fit to ground macro-experience, perhaps some hitherto-unconceived low-level properties could do the job. But again, it's not clear why this appeal to ignorance should have more force than the original Ignorance Objection. As with the subject/subject gap, the low-level/high-level gap seems more-or-less on a par with the structural/phenomenal gap with respect to *prima facie* plausibility. And I am not aware of any considerations that would undermine our *prima facie* justification for accepting a low-level/high-level epistemic gap while leaving the structural/phenomenal gap unscathed.

Another class of epistemic gaps that threaten Russellian monism are those associated with simple phenomenal qualities. As an example, consider the property that we might call “phenomenal yellow”—intuitively, the property of having a sensation of yellow. Our notion of phenomenal yellow seems to be “conceptually primitive” in the sense that no set of truths that do not make explicit reference to phenomenal yellow will *a priori* entail that anything has this specific form of consciousness. Similar epistemic gaps seem to hold for other simple macro-phenomenal qualities, such as phenomenal blue, phenomenal heat, and phenomenal sweetness. Taking phenomenal yellow as our representative example, we can formulate the relevant gap as follows:

Non-yellow/yellow gap: no truth that does not explicitly involve phenomenal yellowness *a priori* entails any truth that explicitly involves phenomenal yellowness.

The notion of “explicit involvement” can be understood by example. Suppose I assert that there is a figure on a plane all of whose points are at a distance of 5 inches from a common point. The proposition I assert may “implicitly” involve the notion of circularity, in that it *a priori* entails the existence of a circle, but it does not explicitly involve circularity. (For fans of structured propositions: the property circularity, or the Fregean sense for circularity, is not a constituent of the proposition.) The intuitive idea behind the non-yellow/yellow gap is that truths about phenomenal yellowness cannot be deduced from a set of base truths unless phenomenal yellowness is explicitly written into the base.⁴

⁴ We might be able to generate exceptions to this principle using tricks analogous to Arthur Prior’s (1960) famous method for deriving normative truths from non-normative truths. For instance, from the fact that (i) x is undergoing a color sensation, and (ii) x is not undergoing sensations $Q_1 \dots Q_n$ (where these are all the color sensations other than phenomenal yellow), it may be possible to deduce that x is having a sensation of yellow. But counterexamples of this kind are irrelevant in the present context. We are concerned with whether truths about phenomenal yellowness could be intelligibly grounded in truths not explicitly about phenomenal yellowness, and facts like (i) and (ii) are not plausible candidates for the grounds of positive macro-experiential truths.

As with the other gaps, the non-yellow/yellow gap has a great deal of intuitive plausibility. Of course, as before, one might resist it through an appeal to ignorance along the lines suggested above, but this line is vulnerable to the same objection as before. There is no reason to find this appeal to ignorance more persuasive than the corresponding response to the original conceivability argument.

The non-yellow/yellow gap, and other gaps of the same kind, create decisive difficulties for panprotopsychoist forms of Russellian monism (at least those motivated by epistemic-gap reasoning). If macro-experiential truths involving phenomenal yellowness can't be *a priori* deduced from truths that do not themselves explicitly involve phenomenal yellowness, then it follows immediately that they cannot be *a priori* deduced from any collection of non-experiential truths, for experiential properties like phenomenal yellowness do not figure in non-experiential truths. If we accept the conceivability-possibility link invoked in the original conceivability argument, we should therefore conclude that panprotopsychoist Russellian monism is false.⁵

The non-yellow/yellow gap also creates nearly decisive difficulties for panpsychoist versions of Russellian monism. Even if fundamental physical entities like quarks and electrons are conscious, surely the conscious lives of quarks and electrons are not so anthropomorphic as to literally involve phenomenal yellowness. After all, the specific phenomenal qualities characterizing human color vision aren't even shared by all other animals with visual systems.

⁵ Some panprotopsychoists hold that experiential properties are relational properties of the form *being acquainted with q*, where *q* is a quality (e.g. yellow) whose instantiation does not by itself entail the existence of experience (Coleman 2015, cf. Chalmers 2015). A proponent of this view might reject the non-yellow/yellow gap as formulated above in favor of the nearby principle that truths involving phenomenal yellowness (i.e. the property of being acquainted with yellow) can't be *a priori* derived from truths that don't explicitly involve *yellowness*. The argument just given does not apply to this package of views, since truths about yellow (as opposed to the property of being acquainted with yellow) need not be experiential truths. Nonetheless, this package of views is vulnerable to an argument closely related to that given in the next paragraph. In particular, even if we allow that fundamental physical entities have "qualities" of some kind, we should not take seriously the hypothesis that quarks and electrons are literally yellow.)

They're certainly not shared by things that don't have anything remotely like a visual system, such as electrons. So if we accept the relevant conceivability-possibility link, the non-yellow/yellow gap leads once again to the conclusion that panpsychist Russellian monism is false.

5. Ghosts and Revelation

In this section, I argue that the Russellian monist's commitment to a strong link between conceivability and possibility makes her vulnerable to another conceivability argument, this time involving the conceivability of *ghosts* rather than zombies. Ghosts are the mirror images of zombies. Zombies share all our physical features but do not have any of our experiential features. Ghosts share all our experiential features but do not have any of our physical features. A ghost, then, is something like a simple Cartesian ego who lacks a body and whose experiences perfectly match those of some actual human being. Ghosts seem to be conceivable. Indeed, the conceivability of ghost scenarios plays a significant role in discussions of external-world skepticism (Cutter 2016: 5). For example, the skeptical scenario Descartes entertains in Meditations I-II, in which "body, shape, extension, movement, and place are all chimeras" (63), is a ghost scenario. It's generally supposed that there is no hidden contradiction in this skeptical hypothesis that would allow one to conclusively rule it out *a priori*. I shall assume, then, that ghosts are conceivable. A bit more specifically, I shall assume it is conceivable that someone should be exactly like me in all experiential respects but lacking entirely in physical parts, or indeed any proper parts whatsoever.

The conceivability of ghosts creates serious difficulties for Russellian monism. To appreciate these difficulties, it will be useful to draw a distinction between *reductive* and *non-reductive* versions of Russellian monism, a distinction which cross-cuts our earlier distinction

between panpsychist and panprotopsychist versions. Reductive Russellian monists hold that macro-experiential properties are identical to complex broadly physical properties (Maxwell 1979, Lockwood 1989, Montero 2010). (Obviously, macro-experiential properties won't be identical to any of the basic or fundamental broadly physical properties, such as the negative charge quiddity. But we can stipulate that the class of broadly physical properties is closed under a restricted class of property-forming operations, e.g. Boolean operations, so that complex properties count as broadly physical if they admit of real definition in terms of the fundamental broadly physical properties and relations via the relevant operations.) By contrast, *non-reductive* Russellian monism holds that macro-experiential properties are grounded in the broadly physical, but not identical to any property definable in broadly physical terms.

The threat to Russellian monism posed by ghosts is most serious for reductive Russellian monism. Consider, for instance, Grover Maxwell's (1979) reductive Russellian monism. Maxwell accepts traditional psycho-neural identities like "pain = C-fiber firing." He also accepts that the terms on each side of these identities are rigid, so they are necessary if true. Of course, such identities *appear* to be contingent, and as Kripke (1980) famously argued, we cannot explain away this appearance of contingency as we might explain away the apparent contingency of other theoretical identities, such as "heat = molecular motion." The apparent contingency in the latter case can be explained by noting that the term on the left-hand side of the identity statement has its reference fixed via a property (causing heat sensations) that is only contingently associated with the term's referent. This strategy doesn't work in the former case, however, since the concept *pain* does not pick out its referent by way of contingently associated properties. Maxwell responds to Kripke, in effect, by saying that the proper analogue of "heat" in "pain = C-fiber firing" is not "pain" but "C-fiber firing." In particular, he maintains that the neuroscientific

concept *C-fiber firing* picks out its referent as the property that occupies a certain structural role, but that neuroscience fails to reveal the complete intrinsic nature of this property.

Let us grant that neuroscience fails to reveal the full nature of C-fiber firing. Still, neuroscience surely tells us something about what an entity must be like in *structural* respects in order to have firing C-fibers. For example, in order to have this property, an entity must be at least partly composed of cells. And even if science does not fully reveal the nature of cell-hood, it is clear that nothing could be composed of cells, and thus nothing could have firing C-fibers, if it did not have proper parts. Since Maxwell identifies the macro-experiential property *being in pain* with C-fiber firing, it follows that nothing can be in pain without having proper parts. The same goes for any broadly physical property that a reductive Russellian monist could reasonably identify with a macro-experiential property like pain. In general, the reductive Russellian monist will presumably want to identify experiential properties with properties of the form *having parts with such-and-such quiddities, arranged in such-and-such way, and interacting in such-and-such way*. Call such properties “quidditative arrangements.” Since plausible forms of reductive Russellian monism will identify experiential properties with quidditative arrangements, and quidditative arrangements cannot be possessed by a simple entity, we may conclude that the following holds for at least some macro-experiential property *E*:

G1. If reductive Russellian monism is true, then it is not possible that something without proper parts has *E*.

From here, we can advance a conceivability argument against reductive Russellian monism:

G2. It is conceivable that something without proper parts has *E*.

G3. If it is conceivable that something without proper parts has *E*, then it is possible that something without proper parts has *E*.

G4. Therefore, reductive Russellian monism is false.⁶

Premise G2 follows from the conceivability of ghosts. Premise G3 takes us from conceivability to possibility. Some will reject this premise, of course, but the Russellian monist who relies on the conceivability argument to rule out standard physicalism is not in a good position to do so.

Taken together with G1, G2 and G3 entail the conclusion G4.

Reductive Russellian monism therefore does not appear to be a stable position, at least if it supposed to be motivated in part by the anti-physicalist conceivability argument. This suggests that the Russellian monist should opt for a *non-reductive* view, according to which macro-experiential properties are not identical to, but merely *grounded in*, quidditative arrangements. However, the non-reductive Russellian monist also faces ghost-related difficulties, even if these difficulties are not quite as decisive as those for reductive Russellian monism. We can bring out the difficulties by considering, and then improving upon, an argument considered (and rejected) by Chalmers (2017: 190), which he calls the “Revelation Argument.” (His formulation targets panpsychist forms of Russellian monism, but it can be easily adapted to apply to panprotopsychist forms as well, as I’ve done below, with modifications in brackets.)

R1. The nature of consciousness is revealed to us in introspection.

R2. If [non-reductive Russellian monism] is correct, consciousness is constituted by a vast array of micro-experiences [or micro-*proto*-experiences].

R3. Whatever constitutes consciousness is part of its nature.

⁶ A similar objection to reductive Russellian monism is discussed in Pautz (ms).

R4. A vast array of micro-experiences [or micro-*proto*-experiences] is not revealed to us in introspection.

R5. Therefore, [non-reductive Russellian monism] is incorrect.

Premise R2 follows from the definition of non-reductive Russellian monism, and R4 is surely beyond reproach, so the key premises are R1 and R3. For our purposes, we can understand the Revelation thesis articulated by R1 in terms of Kit Fine's (1994) notion of an essential truth, where the essential truths about an entity x may be canonically expressed by way of the sentential operator "it lies in the nature of x that ____". In Finean terms, we can understand R1 to express something like the following:

Revelation: for any macro-experiential property Q , if it lies in the nature of Q that P , then one who has Q is in a position to know that P by introspection and *a priori* reflection alone.⁷

Although most physicalists would reject Revelation, it is fairly widely accepted among Russellian monism's proponents and sympathizers (Chalmers 2012; Goff 2017), and anyone who endorses the conceivability argument faces strong pressure to accept it. The most straightforward reason for proponents of the conceivability argument to accept Revelation is the familiar Arnauldian point that if we don't fully grasp the nature of F , there may be necessary connections between F and other properties that aren't transparent to us *a priori*. In that case, the epistemic gap between the (narrowly) physical truths and the experiential truths may just be a symptom of our inadequate grasp of the nature of consciousness rather than an indication of an ontological gap.

⁷ Cf. Byrne and Hilbert's (2007: 77) formulation of Revelation about color.

In response, my opponent might contend that the nature of consciousness is only *partially* revealed to us (Strawson 2006). More specifically, she might contend that enough is revealed to assure us that macro-experience couldn't be grounded in the narrowly physical, but enough is left hidden to leave open the possibility that macro-experience is grounded in the broadly physical. But it's not clear how this delicate balancing act could be motivated (Goff 2017). More importantly, this position seems to be inconsistent with operative assumptions of the conceivability argument. It is clearly inconsistent with the view that conceivability entails possibility, for if some essential truth P about macro-experience is a counterexample to Revelation, then $\sim P$ will be conceivable but impossible.

Granted, there are more sophisticated forms of the conceivability argument, such as Chalmers's (2009) two-dimensionalist conceivability argument, which allow for the conceivability of some impossible propositions, such as the negations of Kripkean *a posteriori* necessities. But even in its sophisticated two-dimensionalist form, the conceivability argument turns out to be inconsistent with the denial of Revelation. Without getting into the technical details of the epistemic two-dimensionalist framework on which the argument relies, let it suffice to note two key assumptions of the argument. The first key assumption is that any necessary truth stated in non-twin-earthable vocabulary is *a priori*, where an expression e as used by a subject S is twin-earthable just in case some possible intrinsic duplicate of S uses e (or a counterpart expression) with a different meaning.⁸ The standard example of a twin-earthable expression is "water," as illustrated by Putnam's (1975) famous twin-earth thought experiment. The second key assumption of the two-dimensionalist conceivability argument is that

⁸ A bit more precisely, the expressions must be "neutral" in the sense that their meanings are independent of actuality. The notion of neutrality is not exactly equivalent to the notion of non-twin-earthability, but they diverge only in *recherché* cases that are irrelevant to our purposes. See Chalmers (2012) for further discussion.

phenomenal expressions like “conscious” are non-twin-earthable. From these two assumptions, Revelation (or something near enough) follows immediately. Proof: Let Q be any macro-experiential property. Suppose it lies in the nature of Q that Q is F , where “ F ” is a non-twin-earthable expression. Then the two assumptions above immediately entail that “ Q is F ” is *a priori*. It therefore should be knowable in principle on the basis of a combination of introspection and *a priori* reflection. (Of course, this only gets us Revelation for essential truths couched in non-twin-earthable terms, but this restricted form of Revelation will suffice for current purposes, since none of the claims to be discussed below make essential use of twin-earthable notions.)

Instead of rejecting Revelation, Chalmers suggests that the best response to the Revelation Argument is to reject R3. He writes,

One can distinguish the nature of a phenomenal property from the grounds (or realizers or constituters) of an instance of that property. It is a familiar point that a single property can be multiply realized by different grounds in different instances, and it is not clear why the same should not also apply to phenomenal properties. It is then coherent to hold that the nature of a phenomenal property is revealed by introspection although the grounds of a specific instance are not. (2017: 190)

Chalmers is certainly correct that full knowledge of the nature of a property needn't put one in a position to know the grounds of a specific instance of it, since that may be a contingent matter. Still, even if it does not belong to the nature of a property that, on a given occasion, it is grounded in, or realized by, such-and-such specific array of properties, the fact that a property is derivative rather than fundamental plausibly *is* something that belongs to its nature. By analogy, one can know what it is to be a table even if one doesn't know that a certain specific instance of tablehood is grounded in a specific configuration of cherry wood. But surely one does not fully know what it is to be a table if one does not know that tablehood is a derivative property, a property that something has in virtue of having certain other features (e.g. a certain shape, a certain set of causal powers, and perhaps a certain causal history). If one doesn't know that

tablehood couldn't be a fundamental property of an object, then one does not fully know what it is to be a table. Likewise, if macro-experiential properties are, like tablehood, metaphysically derivative—properties an individual possesses in virtue of possessing other properties—then anyone with full knowledge of the nature of these properties ought to know this fact about them.

Note the difference between my claim here and premise R3 above. In effect, R3 says that if an instance of F grounds an instance of G, then it lies in the nature of G that an instance of F grounds an instance of G. By contrast, I claim that

R3*. If G is metaphysically derivative, then it lies in the nature of G that G is metaphysically derivative.

We can improve the Revelation Argument by replacing R3 with R3* and replacing R4 with:

R4*. Introspection and *a priori* reflection do not reveal to us that macro-experiential properties are metaphysically derivative.

Let us call the resulting argument the “Revised Revelation Argument.”

Note that Chalmers's point about multiple realizability is no threat to R3*. Indeed, reflection on the multiple realizability of functional properties lends support to R3*. Functional properties typically can be realized by a range of different properties, so an understanding of the nature of a functional property would not tell us how it is realized in a given instance.

Nonetheless, it's plausible that complete knowledge of the nature of a functional property would involve knowledge that it's *realizable*, or even that it's the sort of property that must be realized by some property or other if it is to be instantiated.

The non-reductive Russellian monist might respond to the Revised Revelation Argument by denying R3*, perhaps by maintaining that macro-experiential properties are only *contingently* derivative (Goff 2017: ch. 9). A property *F* is contingently derivative just in case (i) something

has *F* derivatively (i.e. in virtue of having certain other properties), and (ii) possibly, something has *F* fundamentally. Before we evaluate this suggestion, it's worth noting that the non-reductive Russellian monist is more-or-less forced to accept this position, for reasons independent of the (Revised) Revelation Argument. After all, there are conceivable scenarios (e.g. the most straightforward ghost scenarios, as well as property-dualist and substance-dualist scenarios) in which someone has our macro-experiential properties fundamentally. Russellian monists who take conceivability as a guide to possibility should therefore allow that familiar macro-experiential properties *could* be instantiated fundamentally.

Thus it appears that the Russellian monist must—either from a commitment to Revelation or merely from her commitment to a strong conceivability-possibility link—accept that macro-experiential properties are contingently derivative. But there are serious problems with this response. First, it's not clear that there are any contingently derivative properties. In general, derivative properties tend to be *necessarily* derivative. Consider derivative properties like *sitting* or *being a table*. If something is sitting, this is always in virtue of other facts about that thing—the way its limbs are oriented, the way it rests its weight on other things, and so forth. Nor is this a contingent fact about sitting. Sitting simply *could not* be a fundamental feature of a thing. Similarly, if something is a table, this is always in virtue of other facts about that thing—its shape, the way in which it resists penetration, its causal history, and so forth. Being a table *could not* be a fundamental feature of a thing.

It might be said that, while *many* derivative properties are essentially derivative, this is not true of *all* derivative properties, and therefore there's no reason in advance to expect that it must be true of macro-experiential properties. Geoff Lee (2014: 769) mentions mass properties as an example. A composite object might have a mass of 1 gram derivatively, in virtue of being

composed of two particles, each with a mass of 0.5 grams. But it's presumably possible that an elementary particle should have a mass of 1 gram fundamentally.

I don't find this example entirely convincing. It's plausible that the fundamental mass properties—the determinate mass properties that are possessed fundamentally by elementary particles—cannot be possessed by composite things. Of course, we can define a more inclusive notion of mass (“ $mass_i$ ”), a notion that applies to both simple and composite objects, in terms of the more restrictive and more fundamental notion (“ $mass_f$ ”), by defining the $mass_i$ of a composite object to be the sum of the $masses_f$ of its constituent particles. But it's reasonable to suppose that $mass_i$ is never possessed fundamentally, and that the only mass properties that are possessed fundamentally are the (determinate values of) $mass_f$.⁹

Second, and more importantly, even if we allow that there can be contingently derivative properties, it's not clear that this concession would help my opponent respond to the Revised Revelation Argument. Let's say that a property is *weakly* derivative just in case it *could* be instantiated derivatively, and let's say that a property is *strongly* derivative just in case it's weakly derivative and, necessarily, if it's instantiated, then it's instantiated derivatively. The response under consideration says, in effect, that macro-experiential properties are weakly derivative but not strongly derivative. But even the claim that macro-experiential properties are weakly derivative (which any Russellian monist must accept) is enough to create serious difficulties. Given standard assumptions about metaphysical modality, any weakly derivative property will be necessarily weakly derivative. Somewhat more strongly, it's plausible that if a

⁹ It's easy to explain why the derivative notion would be of interest to physics, even if the fundamental laws of nature can be formulated by making reference only to $mass_f$. The key to the explanation is that the laws governing mass ensure that the behavior of a system of particles approximates, in certain respects, the behavior of a hypothetical particle with mass equal to the sum of the masses of the system's constituent particles located at the system's center of gravity. But although the derivative notion of mass is an enormous computational convenience, a Laplacean physicist could do without it.

property is weakly derivative, then it lies in the nature of that property to be weakly derivative. In that case, a commitment to Revelation would entail that one can know by introspection and *a priori* reflection alone that macro-experiential properties are weakly derivative. But this is implausible. It is not clear upon introspection and *a priori* reflection that macro-experiential properties are properties of a sort that even *could* be possessed in virtue of other more basic features. If anything, naïve introspection would seem to suggest that macro-experiential properties enjoy a sort of primitiveness that excludes the possibility of being grounded in anything more basic. Nor will it help my opponent to respond that macro-experiential properties are only *necessarily* weakly derivative, but not (in Fine’s sense) *essentially* weakly derivative. Whether or not this response is plausible in itself, it conflicts with the operative assumptions of the conceivability argument, at least if I am right that it is not *a priori* knowable that macro-experiential properties are weakly derivative. For in that case, this response implies that if *Q* is a macro-experiential property, then “*Q* is weakly derivative” is a necessary but non-*a priori* truth that is formulated entirely in non-twin-earthable vocabulary. For the reasons given above, this would violate not only the crude principle that conceivability entails possibility, but also the more sophisticated principles operative in the two-dimensionalist conceivability argument.

6. Conclusion

We have seen that the Russellian monist occupies an unstable position, at least if she rejects physicalism on the basis of the usual epistemic arguments. To motivate the existence of a “specific” epistemic gap between the (narrowly) physical truths and the experiential truths, we must rely on intuitive judgments about the existence of certain *general* epistemic gaps. But trusting these intuitions leads to the conclusion that there is equally an epistemic gap between the *broadly* physical truths and the experiential truths. If we take conceivability to be a guide to

possibility, as we must if we endorse the conceivability argument against physicalism, we will therefore conclude that Russellian monism is false. Moreover, if we take conceivability to be a guide to possibility, we will conclude that macro-experiential properties could be possessed non-derivatively—for example, by simple disembodied subjects. But as we've seen, the Russellian monist cannot easily accommodate this possibility. The operative assumptions of the conceivability argument therefore turn out to be no more amenable to Russellian monism than they are to physicalism. For those who have abandoned physicalism for the usual reasons, Russellian monism does not offer a comfortable resting place. One must turn back or else march on to dualism. The middle ground is unstable ground.

References

- Alter, T. (2016) "The Structure and Dynamics Argument Against Materialism." *Noûs* 50 (4):794-815.
- Alter, T. & Y. Nagasawa (2012) "What is Russellian Monism?" *Journal of Consciousness Studies* 19 (9-10):67-95.
- Byrne, A. & D. Hilbert (2007) "Color Primitivism." *Erkenntnis* 66(1/2): 73-105.
- Chalmers, D. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- (2002) "Consciousness and its Place in Nature." In D.J. Chalmers (eds.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press.
- (2009) "The Two-Dimensional Argument Against Materialism." In B. McLaughlin & S. Walter (eds.), *Oxford Handbook to the Philosophy of Mind*. Oxford University Press.
- (2012) *Constructing the World*. Oxford: Oxford University Press.

- (2015) “Panpsychism and Panprotopsychism.” In T. Alter & Y. Nagasawa (eds.), *Consciousness in a Physical World: Perspectives on Russellian Monism*. Oxford: Oxford University Press.
- (2017) “The Combination Problem for Panpsychism.” In G. Bruntrup & Ludwig Jaskola, *Panpsychism: Contemporary Perspectives*. Oxford: Oxford University Press.
- Coleman, S. (2015) “Neuro-Cosmology.” In P. Coates & S. Coleman (eds.), *Phenomenal Qualities: Sense Perception, and Consciousness*. Oxford: Oxford University Press, pp. 66-102.
- Cutter, B. (2016) “Color and Shape: A Plea for Equal Treatment.” *Philosophers’ Imprint* 16: 1-11.
- (forthcoming) “Paradise Regained: A Non-reductive Realist Account of the Sensible Qualities.” *Australasian Journal of Philosophy*.
- Descartes, Rene (1998) *Discourse on Method and Meditations on First Philosophy*, 4th edition. D. Cress (trans.). Indianapolis: Hackett.
- Fine, K. (1994) “Essence and Modality.” *Philosophical Perspectives* 8:1-16.
- Goff, P. (2009) “Why Panpsychism Doesn’t Help us Explain Consciousness.” *Dialectica* 63(3): 289-311.
- (2017) *Consciousness and Fundamental Reality*. Oxford: Oxford University Press.
- Jackson, F. (1982) “Epiphenomenal Qualia.” *Philosophical Quarterly* 32: 127-36.
- James, W. (1890/1981) *Principles of Psychology*, vol. 1. Cambridge, MA: Harvard University Press.
- Kripke, S. (1980). *Naming and Necessity*. Oxford: Blackwell.

- Ladyman, J. & D. Ross (2007) *Everything Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.
- Lee, G. (2014) "Unity and Essence in Chalmers' Theory of Consciousness." *Philosophical Studies* 167 (3):763-73.
- Lockwood, M. (1989) *Mind, Brain, and Quantum*. Oxford: Oxford University Press.
- Maxwell, G (1979). "Rigid Designators and Mind-Brain Identity." *Minnesota Studies in the Philosophy of Science* 9: 365-404.
- Montero, B. (2010) "A Russellian Response to the Structural Argument Against Physicalism." *Journal of Consciousness Studies* 17 (3-4):70-83.
- Mumford, S. (2006) "The Ungrounded Argument." *Synthese* 149 (3):471-489.
- Nagasawa, Y. & K. Wager (2016) "Panpsychism and Priority Cosmopsychism." In G. Bruntrup & L. Jaskolla, *Panpsychism: Contemporary Perspectives*. Oxford: Oxford University Press.
- Pautz, D. (ms.) "A Dilemma for Russellian Monists about Consciousness."
- Pereboom, D. (2011) *Consciousness and the Prospects of Physicalism*. Oxford: Oxford University Press.
- Prior, A.N. (1960). "The Autonomy of Ethics." *Australasian Journal of Philosophy* 38 (3):199-206.
- Putnam, H. (1975) "The Meaning of 'Meaning'." *Minnesota Studies in the Philosophy of Science* 7: 131-93.
- Russell, B. (1927) *The Analysis of Matter*. London: Kegan Paul.
- Stoljar, D. (2001) "Two Conceptions of the Physical." *Philosophy and Phenomenological Research* 62 (2):253-81.

————— (2006) *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*. Oxford: Oxford University Press.

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