

Philip Goff

## *Against Constitutive Russellian Monism*

(Forthcoming in Alter, T. & Nagasawa, Y. (Eds.) *Russellian Monism*,  
Oxford University Press)

Russellian monism is a beautiful theory of matter and an attractive solution to the mind-body problem. In the first section of this paper I give a detailed account of what I take Russellian monism to be, how it contrasts with physicalism, and what the motivation is for taking the view very seriously indeed.

In the second section, I turn to the distinction between constitutive and non-constitutive forms of Russellian monism. David Chalmers has argued that constitutive Russellian monism is the most important form of Russellian monism, avoiding disadvantages associated with dualism on the one hand, and standard forms of physicalism on the other.<sup>1</sup> Chalmers does not, however, explore in any detail what exactly the notion of constitution amounts to. Borrowing Theodore Sider's framework for thinking about fundamentality, I attempt to explicate the notion of constitution, and thereby to explicate constitutive forms of Russellian monism.<sup>2</sup> Unfortunately, when fully explicated in this way, constitutive Russellian monism turns out not to be not very plausible, and perhaps incoherent.

The third and final section of the paper brings good news: there is a form of Russellian monism Chalmers does not consider, intelligible emergentist Russellian monism, which remains attractive despite not possessing all the virtues Chalmers attributes to the constitutive view.

### **I – An elegant theory of matter**

#### The problem with pure physicalism

In the public imagination, physics is well on its way to giving a complete description of the fundamental nature of reality. From the scientific revolution onwards, the development of a rigorous experimental method has allowed continuous progress in understanding the nature of space, time and matter. Of course there is a long way to go; physicists have so far been unable to unify our best theory of the very big, i.e. general relativity, with our best theory of the very small, i.e. quantum mechanics. But at some point, it is supposed, these wrinkles will be ironed out and physicists will proudly present the public with the Grand Unified Theory of everything.

Call the view that completed physics will reveal the complete fundamental nature of reality 'pure physicalism.' The troubles with pure physicalism start from the observation that physics, from Galileo onwards, has worked with a very austere vocabulary: its predicates express only mathematical and causal concepts. Think about what physics tells us about an electron. Physics tells us that an electron has negative charge. What does physics have to tell us about negative charge? Rough and ready

---

<sup>1</sup> Chalmers this volume.

<sup>2</sup> Sider 2009, 20012.

answer: things with negative charge repel other things with negative charge and attract other things with positive charge. Physics tells us that an electron has a certain amount of mass. What does physics have to tell us about mass? Rough and ready answer: things with mass attract other things with mass and resist acceleration. All the properties physics ascribes to fundamental particles are characterised in terms of behavioural dispositions. Physics tells us nothing about what an electron *is* beyond what it *does*.

More generally, the information physics provides us with concerning the natural world exclusively concerns its mathematico-causal structure. This is very useful information; it enables us to manipulate nature in all sorts of extraordinary ways, allowing us to build lasers and microwaves, and to fly to the moon. However, if there is more to fundamental reality than can be captured with mathematical and causal concepts, then the description of reality we get from physics must be incomplete. And many philosophers find the idea of a genuine concrete world with a purely mathematico-causal nature unintelligible. Intuitively, wherever there is mathematico-causal structure, there must be some underlying categorical nature that realises that structure; to take a specific case, there must be some categorical nature to an electron that grounds the dispositions physics tells us the electron has. If these intuitions are correct, then no possible world could be completely captured in the language of physics, and hence pure physicalism is necessarily false.

Even if there are genuine possible worlds which can be exhaustively characterised in the language of physics, there is good reason to think that the actual world is not one. For we know with certainty that our world contains phenomenal qualities, qualities such as what it's like to see red, and what it's like to feel pain. And there are powerful arguments, conceivability arguments and the knowledge argument,<sup>3</sup> to the conclusion that a world of pure mathematico-causal structure could not constitute a supervenience base for phenomenal qualities. If these arguments are sound then pure physicalism must be (actually even if not necessarily) false.

There is strong case, then, that physics, for all its virtues, is unable to provide us with a complete description of fundamental reality. Physics provides us with a mathematical description of the fundamental causal workings of the natural world.<sup>4</sup> The formal nature of such a description entails that it necessarily abstracts not only from the reality of consciousness, but from any other real, categorical nature that material entities might happen to have. Just as a mathematical model in economics abstracts away from the concrete features of real world consumers, for example, the nature of their labour and the specific things they buy and sell, so physics abstracts from the concrete features of space, time and matter.

I will assume for the rest of this paper that pure physicalist worlds are indeed incoherent, and that phenomenal qualities do not wholly supervene on mathematico-causal structure. Making these assumptions leaves us with two challenging questions:

1. What is the categorical nature of basic material entities?
2. If physics does not provide us with an adequate supervenience base for the phenomenal qualities we encounter in our experience, how do we explain their instantiation in our world?

---

<sup>3</sup> Chalmers 1996, 2002, Jackson 1982, 1986.

<sup>4</sup> Perhaps we can say more cautiously that physics aims to give a complete description of the causal workings of basic entities in relatively isolated situations. There may be emergent causal powers of which physics remains silent. I discuss causal closure of the microphysical in the third section of the paper.

Russellian monism provides a beautifully unified answer to both of these questions.

### The essence of Russellian monism

According to Russellian monism, the categorical properties of fundamental physical entities – those properties that ground the dispositions physics reveals to us – are phenomenal or protophenomenal properties. I shall say more about the definition of ‘protophenomenal’ presently, but let us work for now with a basic definition according to which they are properties that, although not themselves phenomenal properties, ground phenomenal properties when combined in certain ways.<sup>5</sup> Physics characterises mass in terms of a certain disposition (to resist acceleration and to attract other massive things), but mass itself for the Russellian monist, that is to say, the categorical property that grounds the dispositions expressed by the physical predicate ‘mass’, is a form of consciousness or proto-consciousness.<sup>6</sup> If the categorical properties of basic physical entities are taken to be phenomenal properties, the resulting view is form of panpsychism. If the categorical properties of basic physical entities are taken to be protophenomenal properties, the resulting view is a form of panprotopsyichism.<sup>7</sup>

This single supposition offers hope of a unified answer to both of the questions we ended the last section with. In answer to the first question, the categorical nature of basic material entities turns out to be phenomenal or protophenomenal. In answer to the second question, the presence of this phenomenal or protophenomenal nature is intended to yield a base rich enough to subvene the phenomenal qualities we encounter in our experience. Russellian monists do not in general hold that basic material entities themselves instantiate human phenomenal properties; the consciousness or proto-consciousness of electrons will be of a very basic form. But the crude experience/proto-experience of many basic material entities ground the rich experience of a working human brain.

In its most elegant form Russellian monism is a kind of unorthodox identity theory. Similarly to how physics characterises physical properties in terms of their causal roles, brain science characterises brain states in terms of (i) their causal role in relation to other brains states and behavioural or bodily effects, (ii) their chemical constituents (which are in turn characterised in chemistry in terms of (A) their causal role, (B) their physical constituents). A given brain state, say, c-fibres firing (to continue in the philosophical tradition of using this hackneyed and empirically dubious example), is described by the physical sciences ‘from the outside’, i.e. in terms of the dispositions it grounds and the dispositions of its constituents. But ‘from the inside’ we know that very state as a form of consciousness, say, the feeling of pain. So pain is with identical with c-fibres firing, but the real nature of the state is understood only when it is thought of in phenomenal terms.

A popular argument for reductive physicalism is that it offers a way of way of reconciling platitudes about mental causation with the alleged causal closure of the physical. We want to preserve the commonsense platitude that pain ‘plays the pain role’, i.e. causes avoidance behaviour in response

---

<sup>5</sup> Chalmers 1996: 126-7

<sup>6</sup> Actually, whether this is the right thing to say depends on one’s view about the semantics of the expressions in physics. We may take mass to denote the disposition itself, in which case the phenomenal/protophenomenal base realises mass. For the sake of simplicity, and because it seems to me the more plausible view, I will continue to assume that the expressions of physics rigidly designate categorical properties in terms of the dispositions they ground.

<sup>7</sup> Chalmers 1996/2002/this volume.

to bodily damage. But if the physical is causally closed, then there will be some physical state, c-fibres firing let us say, which plays the pain role. And hence we seem to have too many causes for my screaming and running away when you stick a knife in me. A reductive physicalist does not face the worry of avoidance behaviour being overdetermined, as pain can be identified with c-fibres firing. It is seen to be a great advantage of Russellian monism that, at least in its guise as an unorthodox identity theory, it allows us to say exactly the same thing.<sup>8</sup>

### Distinguishing Russellian monism from impure physicalism

The last section gave us an idea of what Russellian monism is, and why it is an attractive view. However, its exact definition needs refining, as the way I have characterised it thus far does not distinguish it from certain standard forms of physicalism. The problems stem from the basic definition of ‘protophenomenal property’ which I gave at the start of the last section, according to which protophenomenal properties are non-phenomenal properties which ground phenomenal properties when combined in certain ways. The trouble is that any physicalist who believes in phenomenal properties will take phenomenal qualities to be in some sense grounded in more basic properties, and therefore will believe in protophenomenal qualities according to that basic definition.

Even on this definition of ‘protophenomenal’ we can easily distinguish the *pure* physicalist from the Russellian monist, for the latter but not the former believes that basic physical entities have categorical properties. But not all physicalists are pure physicalists. Pure physicalists take the dispositions of fundamental particles to be *brute*, not grounded in some underlying categorical nature. But many physicalists – David Lewis and David Armstrong being two notable examples<sup>9</sup> – believe that the dispositions of fundamental physical entities are grounded in their categorical properties. As physics reveals to us only the dispositional properties of particles, this view entails – as Lewis and Armstrong concede – that there is more to the nature of physical entities than physics can ever reveal to us. Let us call physicalists who are not pure physicalists ‘impure physicalists’.

Alter and Nagasawa have recently defined Russellian monism as the conjunction of the following three claims, in which ‘protophenomenal’ is understood according to the basic definition outlined above:

Structuralism about physics: the basic properties physics describes are structural/relational properties.

Realism about inscrutables: there are inscrutables [i.e. properties which ground the structures/relations physics ascribes], the natures of which are not wholly structural/relational.

(Proto)phenomenal foundationalism: at least some inscrutables are either phenomenal or protophenomenal properties.<sup>10</sup>

---

<sup>8</sup> For some examples of Russellian monism see Russell 1927, Eddington 1928, Feigl 1958/1967, Maxwell 1979, Lockwood, M. 1989, Strawson 1994/2003/2006a, Chalmers 1996, Griffen 1998, Stoljar 2001.

<sup>9</sup> Armstrong 1983, Lewis 2009.

<sup>10</sup> Alter & Nagasawa 2012: 70-1.

According to the Alter/Nagasawa definition, Lewis and Armstrong count as Russellian monists. This is not a welcome result. Russellian monism is understood to be way of incorporating the phenomenon of consciousness into nature *on a non-deflationary* understanding of that phenomenon. Lewis and Armstrong, in contrast, are analytic functionalists about consciousness; the concept of a given conscious state is just the concept of a state that plays a certain causal role.<sup>11</sup>

Perhaps we could just stipulate that Russellian monists reject deflationary analyses of consciousness. However, even with this stipulation, we would be unable to distinguish between Russellian monism and impure versions of the most popular form of contemporary physicalism, the view David Chalmers dubbed 'type-B physicalism'.<sup>12</sup> Type-A physicalists, such as Armstrong and Lewis, take there to be an a priori entailment between completed physics and the phenomenal facts, and usually account for this in terms of some deflationary analysis of consciousness.<sup>13</sup> Type-B physicalists take the phenomenal to supervene on the physical, but deny that there is an a priori entailment from the latter to the former.<sup>14</sup> I will begin by considering the most straightforward versions of Russellian monist and type-B physicalism, according to which there is an a posteriori identity between conscious states and brain states, in order to see starkly the difficulty of distinguishing the two views.

An impure type-B physicalist who believes that the dispositions of fundamental physical entities are grounded in categorical bases will presumably also believe that the dispositions of brain states are grounded in categorical bases, categorical bases which are in turn grounded in the categorical bases of their physical constituents. Just as the categorical nature of an electron is not revealed to us by physics, so the categorical nature of c-fibres firing is not revealed to us by neurophysiology.

So we have the following situation. Both the Russellian monist and the type-B physicalist (in their most simple and elegant forms) take there to be an a posteriori identity between c-fibres firing and pain. Both the Russellian monist and the type-B physicalist take there to be a categorical nature to c-fibres firing, and to each of its fundamental physical constituents, which the physical sciences do not reveal to us. And both the type-B physicalist and the Russellian monist think that at least some of the categorical properties of fundamental physical entities are 'protophenomenal', at least if all this means is that in certain combinations they constitute conscious experience. How then do the two views differ?

I think the views can be distinguished, but we need some terminology in order to do so. Let us say that a concept C denoting a property P is 'transparent' just in case it reveals the nature of P; that is to say, it is possible for someone to come to know a priori, in virtue of possessing C, what it is for P to be instantiated. The concept sphericity is transparent in this sense. For something to be spherical is for it to have all points on its surface equidistant from its centre; if you possess the concept sphericity, and you're clever enough, you can work this out a priori. In contrast we can say that a concept C denoting a property P is 'opaque' just in case it reveals nothing, or very little, of what it is

---

<sup>11</sup> Alter and Nagasawa do go on to say that they are leaving certain commitments of Russellian monism implicit. However, it is highly non-trivial what would need to be added to these three theses in order to exclude the Lewis/Armstrong view.

<sup>12</sup> Chalmers 2002.

<sup>13</sup> For some examples of type-A physicalism, see Armstrong 1968, Lewis 1966/1970/1980/1994, Harman 1990, Dretske 1995, Rey 1995.

<sup>14</sup> For some examples of type-B physicalism, see Loar 1990/2003, Papineau 1993/2002, Lycan 1996, Hill 1997, Hill & McLaughlin 1999, Block and Stalnaker 1999, Perry 2001.

for P to be instantiated. The concept water is an opaque concept. For something to be water is for it to be composed of H<sub>2</sub>O molecules, but you can't work this out a priori simply in virtue of possessing the concepts water and H<sub>2</sub>O.<sup>15</sup>

Finally, we can define a 'transparent rendering' as follows:

*A transparent rendering:* A transparent rendering of description D is a description which is indiscernible from D except that each predicate expressing a non-transparent concept is replaced with a predicate expressing a transparent concept of the same property.<sup>16</sup>

A transparent rendering leaves unchanged facts about reference and extension. It merely, as it were, opens the curtains on each concept, revealing the nature of the property denoted.

Both the Russellian monist and the type-B physicalist take there to be an a posteriori identity between pain and c-fibres firing. But, as I understood them, they disagree as to why that identity is a posteriori. For the Russellian monist, the a posteriority of psycho-physical identities is wholly due to the opacity of the *physical* concepts involved. We cannot know a priori that c-fibres firing is pain, because the concept c-fibres firing denotes its referent 'from the outside' rather than revealing its real nature. If brain science descriptions were transparently rendered, the real phenomenal nature of c-fibres firing would become apparent.

For the type-B physicalist, however, the a posteriority is not due, or at least not solely due, to the opacity of the physical concept. Even if we transparently rendered the concept c-fibres firing, revealing the real categorical nature which underlies the brain science dispositions, there would remain an a posteriori identity between that categorical nature and the phenomenal nature with which it is identical. On the most straightforward understanding of type-B physicalism, the a posteriority of psycho-physical identities is due (in part or in whole) to the opacity of *the phenomenal* concept. It is because the phenomenal concept pain does not reveal the nature of its referent that we cannot know that its referent is brain state X, even when we are conceiving of brain state X under a transparent concept.<sup>17</sup>

There may be scope for a stranger form of type-B physicalism, according to which phenomenal concepts are transparent, and yet psycho-physical identities remain a posteriori even when the relevant brain state is denoted with a transparent concept. Such a form of type-B physicalism would be committed to what I call 'dual revelation': the thesis that, at least in some cases, the nature of a single property can be wholly understood in two conceptually distinct ways. The view would be that I completely understand the nature of pain when I think about it under a phenomenal concept, *and* I completely understand the nature of that same state (but in a radically different way) when I think of

---

<sup>15</sup> Jackson and Chalmers (2001) believe that, given complete physical knowledge (perhaps in conjunction with all phenomenal truths, indexical truths, and the so-called 'that's all' fact) of the actual world one could work out that water is H<sub>2</sub>O, but even they would agree that this cannot be known without empirical information that goes beyond that required to possess the concepts

<sup>16</sup> I develop all these notions in more detail in Goff 2011/Forthcoming a/MS.

<sup>17</sup> Papineau 2006 and McLaughlin 2001 explicitly commit to this form of type-B physicalism. Whilst it is not very common for type-B physicalists to explicitly commit to phenomenal opacity (the thesis that phenomenal concepts are opaque), they tend to defend a semantic externalist account of phenomenal concepts, and this seems to entail phenomenal opacity.

it under a transparent physical concept. It is not obvious that dual revelation is intelligible. Of course I can *refer* to a single property in numerous conceptually distinct ways: by demonstration, by description, in virtue of a causal connection. But it is not clear that sense can be made of my *understanding the nature* of a single property, knowing what it is, say, for something to be spherical, in numerous conceptually distinct ways. Even if intelligible, I have argued elsewhere that dual revelation is deeply problematic.<sup>18</sup>

At any rate, we can distinguish between type-B identity theory and Russellian monist identity theory by saying that the former but not the latter takes psycho-physical identities to be 'robustly a posteriori', in the sense that they remain a posteriori even when the physical side of the identity is transparently rendered.

So much for the distinction between type-B physicalism and Russellian monism in their most simple and elegant forms, i.e. when construed as identity theories. There are forms of both type-B physicalism and Russellian monism which do not identify conscious states with brain states. The type-B physicalist might identify phenomenal properties with physically realised properties, such as functional properties, rather than with specific brain states. The Russellian monist may take c-fibres firing to be realised by, rather than identical with, pain. To achieve a fully general distinction between the two views we can say that the type-B physicalist takes the supervenience of the phenomenal on the physical to be 'robustly a posteriori', in the sense that, even if the complete physical description of reality were transparently rendered, there would fail to be an a priori entailment from the (transparently rendered) physical to the phenomenal.

Russellian monists, in contrast, believe that there is an a priori entailment between a transparent rendering of the physical and the phenomenal. If you could somehow perceive the real categorical nature of my brain states, you would be able to work out the nature of my consciousness experience. If Russellian monism is true then, for a being who could see through the skulls of other people and directly perceive the categorical nature of their brain states, scepticism about other minds would be inconceivable.

This commitment distinguishes Russellian monists from type-B physicalists. However, it does not distinguish them from type-A physicalists, such as Lewis and Armstrong. For Lewis and Armstrong, a transparently rendered description of the physical facts would a priori entail the phenomenal facts in virtue of specifying the pure physical facts (the same facts that a non-transparently rendered description of the physical facts would reveal). I therefore propose to define Russellian monism as the following view:

Russellian Monism There is no a priori entailment between the complete physical description of reality (i.e. the complete description of reality in the vocabulary of fundamental physics) and the phenomenal facts, but there is an a priori entailment between a transparent rendering of the complete physical description of reality and the phenomenal facts.<sup>19</sup>

---

<sup>18</sup> Goff Forthcoming a/MS.

<sup>19</sup> The term 'physical' is due to Galen Strawson (2006). In the formulation of this definition, I am assuming that physical terms like 'mass' rigidly designate categorical properties in terms of the dispositions they ground; on this understanding a transparent rendering of a 'mass' will reveal its underlying categorical nature. If we

We now have the resources to move to a better definition of ‘protophenomenal’ properties, such that we avoid the implication that many type-A and type-B physicalists are committed to such properties. We can say that protophenomenal properties are properties which (i) are not phenomenal properties, (ii) when arranged in certain combinations form a state of affairs S such that there is an a priori entailment between a transparently rendered description of S and some positive states of affairs concerning the instantiation of phenomenal properties, (iii) there is no a priori entailment between a description of the mathematico-causal structure of S and any positive state of affairs concerning the instantiation of phenomenal properties.

One might wonder whether it is important to worry about the distinction between type-B physicalism and Russellian monism.<sup>20</sup> I think it is important for the following reason. There is an important debate over whether the knowledge argument and conceivability arguments are sound, in the sense that they demonstrate some metaphysically significant conclusion. Those philosophers who do not draw metaphysical conclusions from the knowledge/conceivability arguments, such as David Papineau or Michael Tye, tend to be type-B physicalists. Those who do draw metaphysical conclusions from these arguments, such as David Chalmers or Michael Lockwood, tend to be sympathetic to Russellian monism. It is important to clarify the distinction between these two positions in order to see what is at stake in the debate.

## II – Against constitutive Russellian monism

### What is constitution?

I turn now to the form of Russellian monism Chalmers favours: constitutive Russellian monism. Chalmers characterises the view as follows:

*‘Constitutive panpsychism is the thesis that macroexperience [Chalmers’ word for ‘o-experience’]<sup>21</sup> is (wholly or partially) grounded in micro[proto]experience. Alternatively, it is the thesis that macroexperience is constituted by microexperience, or realised by microexperience.’<sup>22</sup>*

He has little to say, however, about what ‘grounding’ or ‘constitution’ or ‘realisation’ are, saying only that on the constitution view, ‘...macrophenomenal truths obtain in virtue of micro[proto]phenomenal truths, in roughly the same sense in which materialists hold that macrophenomenal truths obtain in virtue of microphysical truths.’<sup>23</sup>

---

take ‘mass’ to be a causal role property (see footnote 6), the definition of Russellian monism would go slightly differently. In this case, we could say that the Russellian monist takes there to be an a priori entailment between a complete description of the categorical properties that underlie physical properties (involving transparent concepts of those properties), and the phenomenal facts.

<sup>20</sup> David Papineau and I have a long standing debate over whether there is a significant distinction between the two views. A sample of this debate can be viewed at Goff & Papineau 2012. Our discussions have very much helped me develop the conception of Russellian monism I describe in this section of the paper.

<sup>21</sup> I prefer my term ‘o-experience’, as it leaves open the (rather obscure) possibility that the kind of consciousness we pre-theoretically believe in exists at the micro-level.

<sup>22</sup> Chalmers this volume p. 7.

<sup>23</sup> Chalmers this volume, p. 7.



Let us define the grounding relation as follows: fact X grounds fact Y iff Y obtains in virtue of X obtaining, where I use 'in virtue of' to express the intuitive pre-theoretical notion that something can be the case because something else is the case. Can constitution be characterised wholly in terms of the grounding relation so defined, or perhaps grounding in conjunction with a priori entailment? This would seem to be an inadequate characterisation of constitution, as a grounding relation might obtain in virtue of a relationship of intelligible causation.

Suppose X is the fact that God wills that there be light, and Y is the fact that there is light. Y obtains because X obtains, and there is an a priori entailment from X to Y: if you know that an omnipotent being has willed that there be light then you can work out a priori that there's going to be light. And yet the fact that there is light is clearly not constituted by the fact that God wills that there be light, at least if Y's being constituted by X is supposed to entail that Y is no addition in being to X: in willing that there be light God has created new being *ex nihilo*. To take another example, if the conjunction of dispositional essentialism and determinism is true, then knowing the fundamental nature of reality at a given time could in principle allow us to work out a priori what's going to happen at a later time. And yet this would not entail that the facts about the future are constituted by the facts about the past.

One might object that such causal connections are not instances of grounding. But one would need to move beyond the simple characterisation of grounding I have given above to show that. If grounding is just the in virtue of relation, the relation that holds when one fact obtains because another fact obtains, then causal relationships, e.g. the match lit because it was struck, would seem to constitute instances of grounding.

We might of course use the word 'constitution' in such a way that it does not imply an ontological free lunch, i.e. such that the fact that Y is 'constituted' by X does not entail that Y is nothing over and above X. However, it is clear that Chalmers does want the fact that Y is constituted by X to entail that Y is no addition in being to X, as he takes constituted properties to inherit causal relevance from their grounds. Indeed, this is the principal reason Chalmers takes constitutive forms of Russellian monism to be superior to non-constitutive forms of Russellian monism. According to Chalmers, if the fact that I feel pain, call that fact 'Q', is constituted of some micro-(proto)phenomenal fact M, then M and Q can each be individually sufficient for my pain behaviour, without my pain behaviour being problematically over-determined. As a result, constitutive Russellian monism shares with physicalism the advantage of being able to reconcile mental causation with the causal closure of the micro-physical.

However, this use of constitution to avoid problematic overdetermination works only if constituted facts are no addition in being to their grounds. The fact that the crowd caused the disruption is not overdetermined by the fact that the members of crowd caused the disruption, because the fact that there is a crowd is nothing over and above the fact that there are certain individuals related in certain ways. As a case of contrast, suppose that the members of the crowd are wizards, each with a powerful will akin to the powerful will of God but more limited. It could be that the angry activities bring into existence a demon, intelligibly dependent on, but ontologically additional to, the activities of the wizards. If this demon also causes disruption, we would clearly have a case of overdetermination. Only where we have an ontological free lunch do we avoid worries about causal exclusion.

Let us then define constitution as an in virtue of relation which yields an ontological free lunch: fact X constitutes fact Y iff (i) Y obtains in virtue of X, (ii) Y is nothing over and above X. What is it, though, for a certain fact to be nothing over and above another fact? Reflection on crowds, parties or organisations makes the notion intuitive. But philosophical reflection can render it somewhat mysterious. A crowd is neither identical to its members, nor wholly distinct from them. What is this strange middle way between identity and distinctness? How can fact X involve different objects and properties to fact Y, and yet, from the perspective of serious metaphysics, add nothing beyond the objects of properties already involved in Y? Philosophers trading in ‘nothing over and above’ talk owe us an account of how they get their lunch for free.

The phrase ‘ontological free lunch’ is due to David Armstrong, who claimed that a supervenience relation is sufficient for an ontological free lunch.<sup>24</sup> Chalmers has in the past defended a supervenience based account of physicalism, which suggests that he would acquiesce with Armstrong on this point.<sup>25</sup> However, there is a broad consensus that a mere modal notion like supervenience is not up to the job of explicating the ‘nothing over and above’ relation. Consider the view Terrance Horgan dubbed ‘Moorean Emergentism’:

*Moorean emergentism* – (1) Both phenomenal and physical properties are equally fundamental, (2) phenomenal properties arise from physical properties in virtue of metaphysically necessary psycho-physical laws of nature.<sup>26</sup>

If Moorean emergentism is true, then the phenomenal facts supervene on the physical facts: any possible world W which is a minimal physical duplicate of our world will be a phenomenal duplicate of our world, given that the actual psycho-physical laws (being necessary) also obtain at W. And yet this is clearly not a world where the phenomenal facts are nothing over and above the physical facts. Hence, Moorean emergentism is a counterexample to a wholly supervenience based account of the ontological free lunch.

Although many agree that wholly supervenience-based accounts of physicalism are problematic, there is little consensus as to what they might be replaced with. Physicalism is generally defined as ‘at least’ supervenience, and the details of what more is required often left hazy.<sup>27</sup> Fortunately, in the metaphysics literature in the last five years or so, there has been a great deal of work on the topic of *fundamentality* which can help us make sense of this notion. Perhaps most prominently, Theodore Sider has developed a detailed and systematic framework for thinking about fundamentality.<sup>28</sup> In the next section I outline Sider’s framework; in subsequent sections I will use Sider’s framework to define, and argue against, constitutive Russellian monism.

### Sider on Fundamentality

---

<sup>24</sup> Armstrong 1997: section 2.12.

<sup>25</sup> Chalmers 1996.

<sup>26</sup> Horgan 2006.

<sup>27</sup> Although there have been some interesting proposals, see Horgan 1993, Wilson 1999, Melnyk 2003. I have criticised these proposals elsewhere Goff MS.

<sup>28</sup> Sider 2009/2012.

In 'New work for a theory of universals,' David Lewis argued for the benefits of what we might call 'predicational elitism': distinguishing an 'elite' class of metaphysically privileged predicates that 'carve nature at the joints', from the sprawling slum of metaphysically uninteresting predicates.<sup>29</sup> Thus we might suppose that 'is an electron' carves nature at the joints, whilst 'is grue' or 'is an elephant or an electron' does not. Lewis argues that such a distinction earns its acceptance by its theoretical utility; the distinction allows us to analyse diverse philosophical notions, such as 'intrinsic property' and 'physicalism', and to solve philosophical problems, such as the rule following problem.

Sider develops Lewis's view in two respects. Firstly, where Lewis felt obliged to give some metaphysical account of what it is for a predicate to carve nature at the joints – perhaps all and only the elite predicates express universals – Sider takes the notion of a 'joint carving expression' as primitive. We might clarify this primitive notion in the following way: the world as it is in and of itself<sup>30</sup> has a certain kind of structure, and we represent better to the degree that our conceptual structure mirrors the structure of the world as it is in and of itself. To represent the world perfectly it is not enough to have representations that are true; we must have representations that employ the right concepts.

Secondly, Sider extends Lewis's elitism to linguistic expressions other than predicates. Of most importance for ontology, Sider believes in what we might call 'quantificational elitism', the view that there is a specific meaning of the quantifiers that carves nature at the joints. A candidate meaning of the quantifiers is one that preserves the distinctive inferential role of the quantifiers, for example, preserves the inference from 'Everything is physical', to 'Something is physical.' It is clear that we could invent highly artificial candidate quantifier meanings. Sider gives the following example:

Imagine a person who is logically perfect, maximally opinionated, and totally nuts. His beliefs are logically consistent; for every proposition, he either believes it or believes its negation; and he believes that the moon is made of green cheese, that robots are stealing his luggage, and that Ludwig Wittgenstein was history's greatest philosopher. A candidate meaning on which an arbitrary sentence  $\Phi$  means the same as the English sentence 'according to the (actual) beliefs of the opinionated person,  $\Phi$ ' is then inferentially adequate: the inference rules of quantification theory come out truth-preserving because our opinionated person, being logically perfect and maximally opinionated, believes every logical consequence of everything he believes.<sup>31</sup>

What is being considered above is essentially a silly meaning of the expression 'exists', according to which 'x exists' means 'according to the opinionated person x exists.' Quantifiers with this silly meaning would be inferentially adequate, but they would not carve nature at the joints. The world in and of itself is carved up into certain entities, and the silly quantifier does not range over those entities. Crucially, the quantifier of English, the meaning of which is more influenced by everyday usage than the metaphysical structure of the world, may in a more subtle way also fail to carve

---

<sup>29</sup> Lewis 1983.

<sup>30</sup> Sider does not in fact use the phrase 'the world as it is in and of itself' (as far as I've noticed), but I feel it helps to clarify the notion he's aiming at.

<sup>31</sup> Sider 2009.

nature at the joints. Sider is inclined to think that the world in and of itself is carved up into and only into basic physical entities. But the meaning of the English quantifier may be such that when certain basic physical entities are arranged in certain patterns the English sentence 'There are tables' is true. In this case, the quantifier of English, although perhaps not as silly as the silly quantifier outlined above, does not range over the entities into which the reality is carved in and of itself.

Although most sentences of English will not carve nature at the joints (in the sense that they will be filled with expressions which don't carve nature at the joints), each sentence has, according to Sider, a 'metaphysical truth condition', which states what is required for its truth in a language that perfectly carves nature at the joints (in the sense of containing only expressions that carve nature at the joints). Supposing that the joint carving quantifier ranges only over particles, the metaphysical truth condition of 'There are tables' may be something like 'There are particles arranged table-wise.' I say 'something like' because presumably 'arranged table-wise' is not a predicate that carves nature at the joints. Giving the real metaphysical truth condition for, 'There are tables', would require us to list all the arrangements of particles that would serve to make this sentence true using only expressions that carve nature at the joints.<sup>32</sup> Sider is not optimistic about our capacity to formulate complete metaphysical truth conditions for this or any very interesting sentences of English. However, he takes it to be an important part of a reductive story that we can gesture at 'toy' metaphysical truth conditions that approximate the real metaphysical truth conditions that it is beyond our ken to formulate.<sup>33</sup>

We can distinguish, then, between the fundamental truths – the truths of the perfectly joint carving language – and the truths of English. It may be that there are no fundamental truths about tables, but that 'There are tables' is true in English just in case something like 'There are particles arranged table-wise' is a fundamental truth. If this is so, there is then a clear sense in which the facts about tables are no addition in being to the facts about particles; table facts are not an addition in *fundamental reality*, i.e. the reality which is specified by terms which carve nature at the joints.

In summary, we can define constitution in the Siderian framework as follows:

Truth X constitutes truth Y iff (i) X is a fundamental truth and Y is a non-fundamental truth, (ii) the fundamental reality specified by X satisfies the metaphysical truth condition of Y.

#### Against constitutive panpsychism

At this stage it would be useful to introduce some more terminology. What we really want a theory of consciousness to explain are the phenomenal properties we pre-theoretically associate with humans and other animals. Call facts about the instantiation of these phenomenal properties 'o-phenomenal facts' ('o' for 'ordinary'); call the bearers of these phenomenal properties 'o-subjects.' Call the facts about micro-level subjects/protosubjects in terms of which Russellian monists explain the facts about o-subjects 'micro-phenomenal/protophenomenal facts'; call micro-level subjects/micro-level protosubjects 'micro-subjects'/'micro-protosubjects.'

---

<sup>32</sup> In Sider Forthcoming, he attempts to give a functionalist metaphysical truth condition for the sentence 'Moore has hands.'

<sup>33</sup> Sider 2012: 116-18.

Making sense of protophenomenal versions of constitutive Russellian monism in Sider's framework seems, on the face of it, to be quite straightforward. The constitutive panprotopsychist will say that protophenomenal predicates, but not phenomenal predicates, carve nature at the joints. Reality as it is in and of itself is carved up into particles with protophenomenal properties, but not into subjects of experience. Sentences quantifying over subjects of experience have metaphysical truth conditions that involve quantification over fundamental particles and predication over protophenomenal properties. If all this works out, there would then be a clear sense in which the o-phenomenal facts are no addition in being to the protophenomenal facts.

Interestingly, it seems that panpsychist versions of constitutive Russellian monism cannot be coherently specified using Sider's framework. The problem is that the kind term 'subject of experience' is either a joint carving expression or it is not. The fundamental truths, for the constitutive panpsychist, concern micro-level subjects of experience. All the expressions in a fundamental truth are joint carving, and hence the kind term 'subject of experience', as it appears in fundamental micro-phenomenal truths, must be joint carving. But if the kind term 'subject of experience' as it appears in certain micro-phenomenal truths is joint carving, then the kind term 'subject of experience' must also be joint carving as it appears in truths concerning o-subjects of experience, such as you and I. This follows from the sub-propositional focus of Sider's framework. Whether or not a certain expression is joint carving is not dependent on whether the truth it is contained in is a fundamental truth; rather whether or not a given truth is fundamental depends on whether or not it is formed from joint carving expressions.

We can put the argument as follows:

1. If panpsychism is true, then at least some tokens of the kind term 'subject of experience' are joint carving (i.e. those used to specify the fundamental micro-phenomenal truths).
2. If some tokens of the kind term 'subject of experience' are joint carving, then all tokens of the kind term 'subject of experience' are joint carving, including those contained in o-phenomenal truths.
3. If tokens of the kind term 'subject of experience' that are contained in o-phenomenal truths are joint carving, then macro-level subjects of experience are fundamental entities, i.e. entities which are quantified over in fundamental truths.
4. Therefore, if panpsychism is true, macro-level subjects of experience are fundamental entities.

In Sider's framework, panpsychism entails that o-subjects and micro-subjects are fundamental entities. It is, therefore, inconsistent with constitutive Russellian monism: if o-subjects are fundamental entities in their own right, then they cannot be constituted by micro-level goings on.

A bundle theorist constitutive panpsychist might respond by denying that the kind term 'subject of experience' is used in characterising the fundamental micro-phenomenal truths, claiming that the fundamental micro-phenomenal truths involve bundles of micro-phenomenal properties. However, when considering such a view, we could simply swap the property kind term 'subject of experience' in the above argument for the property kind term 'phenomenal property'. We would then reach the conclusion that o-phenomenal properties are fundamental entities, and again end up with fundamental entities involved in o-phenomenal facts. Given that such entities are fundamental, they

cannot be constituted by micro-level goings on, which is inconsistent with constitutive Russellian monism.<sup>34</sup>

The constitutive panpsychist might try to claim that the joint carving expressions are not the determinables 'subject of experience' or 'phenomenal property', but terms denoting the specific determinate states of consciousness possessed by micro-subjects, states of consciousness which o-subjects do not possess. However, the panpsychist can only articulate what she takes the fundamental truths to using phenomenal concepts that we already have, and those concepts are drawn from our understanding of o-phenomenology. To say that particles have some weird and whacky properties that we don't have a grip on is not to commit to panpsychism. The pansychist's view of the fundamental truths is that they involve 'consciousness', that there is 'something that it is like' to be a fundamental particle. And these expressions which she takes to be involved in articulating the fundamental truths also apply to us. Once we allow that at least some of our ordinary expressions pertaining to consciousness are joint carving, it is hard to see how we could avoid there being at least some fundamental o-phenomenal truths.

I have been assuming that the panpsychist thinks truths about consciousness are amongst the fundamental truths. There could be a strange view according to which consciousness is ubiquitous at the micro-level, but there are no fundamental truths about consciousness. For example, it might be the case that the English sentence 'Particles are conscious' is true, but that this truth has a metaphysical truth condition which does not involve phenomenal predicates. I think this kind of view is best seen as an odd kind of panprotopsychism (or perhaps even type-B physicalism). In metaphysics we are ultimately interested in fundamental truth, and so a view that doesn't take fundamental truths to involve consciousness isn't appropriately classed as a form of panpsychism. In any case, the substantive point is that we have ruled out the following option: being metaphysically serious about micro-consciousness whilst being metaphysically lightweight about o-consciousness. Either 'consciousness' carves nature at the joints or it doesn't; if there are fundamental micro-phenomenal truths then there are fundamental o-phenomenal truths.

Of course constitutive panpsychists might try to find some framework other than Sider's for making sense of their view, and so I cannot conclusively rule out panpsychist forms of constitutive Russellian monism.<sup>35</sup> But to my mind what this has brought out is that you're either metaphysically serious about consciousness or you're not. It doesn't really make sense to be metaphysically serious about some subjects (at the micro-level) and metaphysically lightweight about other subjects (o-subjects). Panpsychism entails that o-subjects are fundamental.

#### Against all forms of constitutive Russellian monism

In the last section I argued that panpsychism is an essentially layered view of reality, and hence that we cannot make sense of panpsychist forms of constitutive Russellian monism. One might think that this gives the advantage to panprotopsychism over panpsychism. However, I will now try to show that, at least when explicated in the Siderian framework, all forms of constitutive Russellian monism

---

<sup>34</sup> The panpsychist may be a nominalist, refusing to quantify over properties but taking phenomenal predicates to carve nature at the joints. Such a view would face essentially the same problem of ending up with fundamental o-phenomenal truths.

<sup>35</sup> Fine 2001 offers a propositional rather than a sub-propositional account of metaphysically heavyweight facts. Sider argues for the superiority of his sub-propositional account in his 2012: 8.3.

are deeply implausible, and perhaps incoherent. (In what follows I will equate 'constitutive Russellian monism' with constitutive Russellian monism explicated in Sider's framework.)

If constitutive Russellian monism is true, then truths concerning o-subjects, for example, 'There is something that feels pain', have metaphysical truth conditions that quantify over particles, either non-conscious particles, or particles with very different phenomenal properties to those expressed by the predicates of the original English sentence. As I said above, Sider thinks that it would be unreasonable to expect us to be able to give a precise statement of the metaphysical truth condition of any interesting ordinary language sentence, but that it is important for a reductive view of X to be able to gesture at a rough 'toy' approximation of the metaphysical truth conditions of truths about X. We can't state the complete disjunction of particle arrangements necessary for the truth of 'There are tables', but we have a (very) rough idea about what facts about particles must obtain in order for that sentence to be true.

Sider is surely right about this. In order for it to be plausible that there is a metaphysical truth condition of 'Bill is in pain' concerning a large number of particles, we must be able to gesture at it. The only way I can see that such a gesture might be given is by offering some kind of deflationary analysis of consciousness, i.e. analytic functionalism, analytic behaviourism, or analytic information-theoretic representationalism. Suppose we accept David Lewis's view that for something to be in pain is for it to have an inner state that plays a certain causal role mediating between bodily damage and avoidance behaviour. We would then have a rough idea of how particles must be arranged for 'Bill is in pain' to be true, akin to the rough idea we have of how particles must be arranged for 'There is a table' to be true.

However, Russellian monists reject such deflationary analyses of phenomenal truths. On a deflationary analysis of phenomenal truths, they are a priori entailed by the complete physical truth, whilst it is part of Russellian monism as I have defined it that o-phenomenal truths are not entailed by the complete physical truth. And this is not an idiosyncratic feature of my definition: Russellian monism is an attempt to find a place for the phenomenon of consciousness on a non-deflationary understanding of that phenomenon.

What is required of reality as it is in and of itself for there to be something in pain? In the absence of a deflationary analysis, it's hard to see what metaphysical truth condition we can give for this sentence other than one that is structurally isomorphic with the original English sentence, i.e. one that employs the same predicates and quantifies over the same entities. And if the metaphysical truth condition of 'There is something that feels pain' is just 'There is something that feels pain', then 'There is something that feels pain' is a fundamental truth and subjects that feel pain are fundamental entities.

(This needs to be qualified slightly. Obviously some phenomenal predicates will not carve nature at the joints, e.g. grue-like phenomenal predicates. In what follows I will take phenomenal truths to be truths that involve only quantification and non-disjunctive/conjunctive predicates expressing direct phenomenal concepts, i.e. concepts which are formed wholly on the basis of attending to a phenomenal quality in introspection.)<sup>36</sup>

---

<sup>36</sup> The notion of a direct phenomenal concept comes from Chalmers 2003.

I think the type-B physicalist would have a good response to this concern. Most type-B physicalists defend a semantic externalist account of phenomenal concepts, whereby the extension of the concept is fixed by facts outside of what is a priori accessible, such as causal or teleological facts. As a consequence, in contrast to, say, the concept water, phenomenal concepts lack descriptive content. Many type-B physicalists appeal to this difference to explain why there is an explanatory gap between the physical and the phenomenal facts that is lacking between the physical facts and the facts about water. It is plausible that the concept water connotes the description 'the colourless, odourless, stuff in oceans and lakes,' and that if we know enough about the physical facts we can know that H<sub>2</sub>O satisfies that description. In contrast, the concept pain lacks any such associated description, which is why we can't move a priori from knowing that someone's c-fibres are firing to knowing that they are in pain.<sup>37</sup>

The type-B physicalist might make a similar response to a Siderian demand to know the metaphysical truth conditions of phenomenal sentences:

The concepts water and table refer by description, which is why the metaphysical truth conditions of truths about water and tables are a priori accessible (which is not to say the complete metaphysical truth condition is easy to access, but a toy approximation of it is easily accessible). But the extension of phenomenal concepts is determined by facts which are not a priori accessible, perhaps causal or teleological facts, and hence the metaphysical truth conditions of phenomenal truths will not be a priori accessible. The metaphysical truth condition of 'There is something that feels pain', is 'There is something with particles arranged c-fibre firing-wise', but because the metaphysical truth condition is determined by facts outside of what is a priori accessible, armchair reflection will never reveal this.

This response could be seen to constitute a kind of 'phenomenal concept strategy' response to the apparent absence of reductive metaphysical truth conditions for phenomenal truths.<sup>38</sup>

However, if the metaphysical truth conditions of o-phenomenal truths are not a priori accessible (because phenomenal concepts are opaque), then the necessitation of phenomenal truths by physical truths (if such a thing there be) will be robustly a posteriori. If it is not a priori what is fundamentally required for 'There is something that feels pain' to be true, then there can be no a priori connection between the fundamental facts and the truth of 'There is something that feels pain.' But constitutive Russellian monism is defined such that there is an a priori entailment between the fundamental truths and the o-phenomenal truths (at least if we assume, as I will, that the constitutive Russellian monist takes the fundamental truths to be given by a transparent rendering of a complete physical description of reality).

Let's bring everything together. The definition of Russellian monism has a negative component – there is no a priori entailment from the complete physical description of reality to the o-phenomenal facts – and a positive component – there is an a priori entailment from a transparent

---

<sup>37</sup> See Papineau 1998 for a clear expression of this kind of view.

<sup>38</sup> The term 'phenomenal concept strategy' comes from Stoljar 2005.



rendering of the complete physical description of reality to the o-phenomenal facts. The negative component of the definition supports the following premise: if the metaphysical truth conditions of phenomenal truths are a priori accessible then each phenomenal truth is structurally isomorphic with its metaphysical truth condition (The reason the negative component supports this premise is that only if we accept deflationary analyses is it plausible to suppose that a phenomenal truth has an a priori accessible metaphysical truth condition with which it is not structurally isomorphic, and the negative component rules out deflationary analyses). The positive component of the definition supports the following premise: the metaphysical truth conditions of phenomenal truths are a priori accessible.

We can present, then, the following argument against all forms of constitutive Russellian monism:

1. If constitutive Russellian monism is true, then there is no a priori entailment between the complete physical description of reality and the o-phenomenal truths, but there is an a priori entailment between the fundamental truths (i.e. a transparent rendering of the complete physical description of reality) and the o-phenomenal truths.
2. If there is no a priori entailment between the complete physical description of reality and the o-phenomenal truths, then either the metaphysical truth conditions of phenomenal truths are not a priori accessible, or each o-phenomenal truth is structurally isomorphic with its metaphysical truth condition and hence o-phenomenal truths are fundamental truths.
3. If there is an a priori entailment between the fundamental truths and the o-phenomenal truths, then the metaphysical truth conditions of o-phenomenal truths are a priori accessible.
4. Therefore, if constitutive Russellian monism is true, then o-phenomenal truths are fundamental truths.
5. If constitutive Russellian monism is true, then it is not the case that o-phenomenal truths are fundamental truths.
6. Therefore, constitutive Russellian monism is false.

For those who prefer symbols:

P=the complete physical description of reality

TP=a transparent rendering of the complete physical description reality

Q=an arbitrary o-phenomenal truth

1. Constitutive Russellian monism  $\rightarrow$  (P $\rightarrow$ Q is not a priori) & (TP $\rightarrow$ Q is a priori).
2. (P $\rightarrow$ Q is not a priori)  $\rightarrow$  ( $\sim$ (The metaphysical truth condition of Q is a priori)  $\vee$  (Q is a fundamental truth))
3. (TP $\rightarrow$ Q is a priori)  $\rightarrow$  The metaphysical truth condition of Q is a priori.
4. Therefore: constitutive Russellian monism  $\rightarrow$  Q is fundamental truth.
5. Constitutive Russellian monism  $\rightarrow$   $\sim$ (Q is a fundamental truth)
6. Therefore: constitutive Russellian monism is false.

Of course this argument depends on explicating constitutive Russellian monism in Sider's framework. However, the onus is on the proponent of constitutive Russellian monism to offer an account of constitution which avoids these concerns.

### The intuitive moral of the story

What I have offered above is a fairly complex argument, employing a fairly complex metametaphysical framework. But I take this to be an attempt to rigorously explicate a fairly intuitive thought. If you want to make sense of facts about o-consciousness being constituted of facts not involving o-consciousness, then you have to offer hope that there is some kind of *reductive analysis* of o-consciousness: some way of understanding the o-phenomenal facts in more fundamental terms.<sup>39</sup> You might go the way of the type-A physicalists and give a deflationary analysis of phenomenal truths. Or you might go the way of the type-B physicalists and hold that we have no a priori access to the nature of consciousness, which opens us up to the possibility of discovering empirically that the real nature of consciousness can be grasped in non-experiential terms. The Russellian monist rejects both of these options, and thus is left with no alternative but to take o-phenomenal truths to be fundamental.

Couldn't the constitutive Russellian monist claim that we know *something but not everything* about the nature of phenomenal truths? This is the response one often hears from Russellian monists in conversation, and has been suggested in print by Galen Strawson<sup>40</sup> (in response to me<sup>41</sup>) and Derk Pereboom.<sup>42</sup> The problem is that we can now just shift the argument to focusing on *the bit of consciousness we have transparent access to*, and argue that that property must be fundamental, given that we have no way of understanding it in more fundamental terms.

In any case, constitutive Russellian monists often gesture towards the view that phenomenal concepts are *translucent*, revealing something but not everything of the states they denote, but have never as far as I'm aware worked out the details. Suppose I form a direct phenomenal concept of the phenomenal red in my experiential field right now. What is the aspect of phenomenal red which is transparently revealed in my conception? What is the aspect of phenomenal red which is opaquely denoted in my conception? (In the context of the above argument, the challenge will be to distinguish the bit of the metaphysical truth condition which is a priori accessible from the bit which is not). This is not a matter of telling us the nature of protophenomenal properties, many Russellian monists takes these to be permanently beyond our grasp; rather my request is for a detailed account of phenomenal *concepts* such that they turn out to be translucent.<sup>43</sup> If the constitutive Russellian monist wants to appeal to translucency she is obliged to fill in the details.

### **III – Intelligible emergentism**

Chalmers contrasts constitutive Russellian monism with *emergentist* Russellian monism: the view that o-experience arises from micro-(proto)experience in virtue of causation rather than constitution. Chalmers describes a kind of brute emergentism, according to which o-experience is strongly emergent from micro-(proto)experience, i.e. o-experience arises from micro-(proto)experience but

---

<sup>39</sup> I don't use the term 'reductive analysis' such that a reductive analysis is by definition a priori.

<sup>40</sup> Strawson 2006b.

<sup>41</sup> Goff 2006.

<sup>42</sup> Pereboom 2011: 114-5/this volume.

<sup>43</sup> See Schroer 2010 for a very good physicalist attempt to make sense of phenomenal translucency. I expect the Russellian monist who wants to appeal to phenomenal transparency to give a similarly detailed account of how phenomenal concepts work such that they are translucent.

without its emergence being predicable from knowledge of the micro-(proto)phenomenal facts alone:

One sort of emergentist panpsychism might hold that there are contingent laws of nature that specify when certain microexperiences give rise to certain macroexperiences [Chalmers' word for 'o-experience']. Another might hold that there are laws of nature connecting microphenomenal properties and macrophysical properties to macrophenomenal properties, without there being any constitutive connection between microphenomenal and macrophenomenal.<sup>44</sup>

Thus, the view Chalmers uses as a contrast to constitutive Russellian monism is a kind of pan(proto)psychist version of British emergentism.<sup>45</sup> However, it is not clear that brute panprotopsychism is coherent, given that protophenomenal properties by definition a priori entail phenomenal properties.<sup>46</sup> If the relation between micro-protosperience and o-experience is causal, then the causation must be intelligible rather than brute. Near the beginning of section II we considered divine creation and dispositional essentialism as examples of intelligible causation. Analogously, it could be that the o-phenomenal facts are intelligibly caused by the micro-protophenomenal facts. Although not conceptually obligatory, we might think of the relationship between micro-phenomenal and o-phenomenal facts in the same way.<sup>47</sup>

Let us call the following position 'intelligible emergentism':

- (i) o-phenomenal facts are a priori entailed by micro-(proto)phenomenal facts,
- (ii) Both micro-(proto)phenomenal and o-phenomenal facts are fundamental, and hence the grounding relationship between these two kinds of fact is causal rather than constitutive.<sup>48</sup>

Combining intelligible emergentism and Russellian monism we get the following view:

- (iii) Intelligible emergentist Russellian Monism (i) There is no a priori entailment between the complete physical description of reality (i.e. the complete description of reality in the vocabulary of fundamental physics) and the phenomenal facts, but there is an a priori entailment between a transparent rendering of the complete physical description of reality and the o-phenomenal

---

<sup>44</sup> Chalmers this volume: 7.

<sup>45</sup> See Mill (1843), Brain (1870), Lewes (1875), Alexander (1920), Morgan (1923), Broad (1925). For a very good discussion of British emergentism, see McLaughlin 1992.

<sup>46</sup> Chalmers (this volume) also defines 'protophenomenal' such that protophenomenal properties a priori entail phenomenal properties.

<sup>47</sup> If we define Russellian monism, as I have done, as involving a priori entailment, then it is conceptually obligatory to think of a Russellian version of emergentist panpsychism as involving intelligible rather than brute causation.

<sup>48</sup> I think of causal relationships as non-constitutive grounding relations, i.e. grounding relations between distinct and fundamental truths.

facts, (ii) both micro-(proto)phenomenal and o-phenomenal facts are fundamental, and hence the grounding relationship between these two kinds of fact is causal rather than constitutive.

There are strong arguments associated with the ‘combination problem’ to the conclusion that there could be no a priori entailment from the micro-phenomenal to the o-phenomenal.<sup>49</sup> It seems on the face of it that we can conceive of micro-phenomenal facts obtaining in the absence of any o-phenomenal facts, which is inconsistent with there being an a priori entailment from the former to the latter. I have defended elsewhere a solution to this difficulty involving the notion of *phenomenal bonding*.<sup>50</sup> If there is a solution to the combination problem, and we can make sense of an intelligible connection between micro-(proto)phenomenology and o-phenomenology, then intelligible emergentist Russellian monism provides an elegant place for o-consciousness in nature, much more elegant than standard emergentism with its reliance of brute laws.<sup>51</sup>

What about causal exclusion worries? Emergentist forms of Russellian monism do not seem adequately equipped to reconcile the causal efficacy of o-consciousness with *micro*-physical causal closure. However, they do have the resources to assuage exclusion concerns which come from a commitment to *macro*-physical causal closure. Suppose one thought that the event of c-fibres firing in my brain was sufficient cause for my screaming and running away after a knife had been stuck in me.<sup>52</sup> A non-Russellian property dualist might worry that this crowds out the possibility that my feeling pain caused me to scream and run away, and may consequently resort to epiphenomenalism. A Russellian monist, however, avoids these concerns. Either c-fibres firing just is (in its real essential nature) my feeling of pain, or my feeling of pain constitutes c-fibres firing (in the latter case c-fibres firing is a purely dispositional property realised by my o-experience).

In this way, the emergentist Russellian monist takes the macro-physical to be nothing over and above the macro-phenomenal, and hence there is no worry about causal competition between mental and physical at the level of brains. This is a not insignificant advantage of the view. However, any emergentist takes o-experience to be something over and above micro going on, both phenomenal and non-phenomenal. It must be conceded that if the micro-level is causally closed, this would seem to crowd out the causal efficacy of the macro-level.

How satisfactory intelligible emergentist Russellian monism is, then, may depend on the force of the arguments for micro-physical causal closure. I don’t have space to pursue the matter in much detail here, but for my own part I remain unpersuaded that there is a strong argument for microphysical causal closure. Often causal closure of the physical is just assumed without argument. And where

---

<sup>49</sup> Goff 2009, discussed in Chalmers this volume.

<sup>50</sup> Goff Forthcoming b.

<sup>51</sup> Goff Forthcoming b. The argument I present in this paper might be seen as a way of pressing a version of the combination problem against constitutive Russellian monism. Interpreted in this way, as a challenge to constitutive pan(proto)psychism, I take the combination problem to be insoluble. Interpreted as a challenge to the more general view that higher-level phenomenal facts can be intelligibly grounded in more basic (proto)phenomenal facts, I take the combination problem to be soluble. I talk about these two ways of understanding the combination problem in much more detail in Goff MS.

<sup>52</sup> I have very much benefited from many discussions on this topic with Hedda Hassel, who provides a rich discussion of these issues in Hassel forthcoming.

there is an argument it is often unclear whether that argument supports *microphysical* causal closure, or the causal closure of the physical more generally construed, the latter view being consistent with an indispensable role for brain-level events.<sup>53</sup> Until the case for micro-physical causal closure is clarified, the status of intelligible emergentist Russellian monism will also remain unclear.

## Conclusion

At this stage of enquiry, intelligible emergentist Russellian monism is a view we should take seriously. For those who accept the soundness of the knowledge argument or the conceivability arguments, neither pure nor impure physicalism can be true, as they are both inconsistent with the one natural phenomenon we know with certainty to exist: consciousness. Of course, there are a range of anti-physicalist alternatives. However, even as a layered view of the world, Russellian monism looks at present to be the most elegant way of finding a home for consciousness.<sup>54</sup>

## References

- Alter, T. & Nagasawa, Y. (2012) What is Russellian monism? *Journal of Consciousness Studies* 19: 9-10, 67-95.
- Alexander, S. (1920) *Space, Time, and Deity*, 2 Volumes, London: Macmillan.
- Armstrong, D. (1968) *A Materialist Theory of Mind*, London: Routledge and Kegan Paul.
- Armstrong, D. (1983) *What is a Law of Nature?* Cambridge: Cambridge University Press.
- Armstrong, D. (1997) *A World of States of Affairs*, Cambridge: Cambridge University Press.
- Bain, A. (1870) *Logic*, Books II and III, London: Longmans, Green, Reader & Dyer.
- Block, N. and R. Stalnaker (1999) Conceptual Analysis, Dualism and the Explanatory Gap, *Philosophical Review* 108/1: 1–46.
- Broad, C. D. (1925) *The Mind and Its Place in Nature*, London: Routledge and Kegan Paul.
- Chalmers, D. J. (1996) *The Conscious Mind: Towards a Fundamental Theory*, Oxford: Oxford University Press.
- Chalmers, D. J. (2002) Consciousness and its Place in Nature, in *Philosophy of Mind: Classical and Contemporary Readings*, ed. D. J. Chalmers, Oxford, New York: Oxford University Press, 247-72.

---

<sup>53</sup> See the appendix of Papineau 2002 for an extended argument for the causal closure of the physical.

<sup>54</sup> I am very grateful for fantastic comments from Torin Alter, Luke Roefls, Hedda Hassel and David Chalmers. I gave early versions of this paper at the 'Towards a non-physicalist solution to the mystery of consciousness' workshop organised by Yujin Nagasawa and Max Velmans, and at the 'Metaphysics and Ontology of Phenomenal Qualities' conference, which was part of the 'Phenomenal Qualities' AHRC project. I very much benefitted from comments from participants at both events. I am also grateful for the precious research time afforded to me by my post-doctoral position as part of the 'Phenomenal Qualities' project, and my post-doctoral position at the 'Centre for Consciousness' at the Australian National University.

- Chalmers, D. J. (2003) The content and epistemology of phenomenal belief, in Q. Smith & A. Jolic (Eds.) *Consciousness: New Philosophical Perspectives*, Oxford University Press.
- Chalmers, D. J. (this volume) 'Panpsychism and panprotopsyism.'
- Dretske, F. (1995) *Naturalizing the Mind*, Cambridge, MA: MIT Press.
- Eddington, A. (1928) *The Nature of the Physical World*, New York: Macmillan.
- Feigl, H. (1967) *The "mental" and the "physical": the Essay and a Postscript*, Minneapolis: University of Minnesota Press.
- Fine, K. (2001) The question of realism, *Philosophers Imprint*, 1: 1-30.
- Goff, P. (2006) 'Experiences don't sum,' *Journal of Consciousness Studies*, 13.
- Goff, P. (2009) Why pansychism doesn't help explain consciousness, *Dialectica*, 63.
- Goff, P. (2011) A posteriori physicalists get our phenomenal concepts wrong, *Australasian Journal of Philosophy* 89: 2, 191–209.
- Goff, P. (Forthcoming a) 'Real acquaintance and the phenomenal concept strategy,' in S. Coates & S. Coleman (Eds.) *Phenomenal Qualities*, Oxford: Oxford University Press.
- Goff, P. (Forthcoming b) The phenomenal bonding solution to the combination problem, in G. Bruntrup & L. Jaskolla (Eds.) *Panpsychism*, Oxford University Press.
- Goff, P. (MS) *Consciousness and Fundamental Reality*.
- Goff, P & Papineau, D. (2012) 'On physicalism,' *Philosophy TV*, <http://www.philostv.com/?s=papineau&submit.x=-1175&submit.y=-224&submit=Search>
- Griffen, D. R. (1998) *Unsnarling the World-Knot: Consciousness, Freedom, and the Mind-body Problem*, Minneapolis: University of Minnesota Press
- Harman, G. (1990) The Intrinsic Quality of Experience, *Philosophical Perspectives* 4: 31–52.
- Hassel, H. Forthcoming. *Panpsychism and Causation: A New Argument and a Solution to the Combination Problem*, PhD thesis, University of Oslo.
- Hill, C. (1997) Imaginability, Conceivability, Possibility and the Mind-Body Problem, *Philosophical Studies* 87:1, 61–85.
- Hill, C. & B. McLaughlin (1999) There are Fewer Things in Reality than are Dreamt of in Chalmers's  
*Philosophy, Philosophy and Phenomenological Research* 59:2, 445–54.
- Horgan, T. (1993) From supervenience to superdupervenience: Meeting the demands of a material world, *Mind* 102, 555-86.

- Jackson, F. 1982. 'Epiphenomenal Qualia,' *Philosophical Quarterly* 32.
- Jackson, F., 1986. 'What Mary Didn't Know,' *Journal of Philosophy* 83.
- Levine, J. (1983) Materialism and Qualia: the Explanatory Gap,' *Pacific Philosophical Quarterly* 64: 9, 354–61.
- Lewes, G. H. 1875: *Problems of Life and Mind*, Volume 2. London: Kegan Paul, Trench, Turbner, & Co.
- Lewis, D. (1966) An Argument for the Identity Theory, *Journal of Philosophy* 63/1: 17–25.
- Lewis, D. (1970) How to Define Theoretical Terms, *Journal of Philosophy* 67/13: 427–46.
- Lewis, D. (1980) Mad Pain and Martian Pain, in N. Block (Ed.) *Readings in the Philosophy of Psychology, Vol. I*, (Ed.) N. Block, Harvard: Harvard University Press: 216–22.
- Lewis, D. (1983) New work for a theory of universals, *Australasian Journal of Philosophy* 61, 343–77.
- Lewis, D. (1994) Reduction of Mind, in *Companion to the Philosophy of Mind*, in S. Guttenplan (Ed.), Oxford: Blackwell: 412–31.
- Lewis, D. (2009) Ramseyan Humility, in D. Braddon-Mitchell & R. Nolan (Eds.) *Conceptual Analysis and Philosophical Naturalism*, Cambridge, MA: MIT Press.
- Loar, B. (1990) Phenomenal States, in *Philosophical Perspectives* 4: 81–108.
- Loar, B. (2003) Qualia, Properties, Modality, *Philosophical Issues* 13/1: 113–29.
- Lockwood, M. (1989) *Mind, Brain and the Quantum*, Oxford: Oxford University Press.
- Lycan, W. G. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.
- Maxwell, G. (1979) Rigid designators and mind-brain identity, in C. W. Savage (Ed.) *Perception and Cognition*, Minneapolis: University of Minnesota Press, 365–404.
- McLaughlin, B. 1992: "The Rise and Fall of British Emergentism", in A. Beckermann, H. Flohr & J. Kim (Eds.) 1992: *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*, Berlin: Walter de Gruyter, 49-93.
- McLaughlin, B. P. (2001) A defense of new wave materialism: A response to Horgan and Tienson, in Gillett and Loewer (2001).
- Melnyk, A. (2003) *A Physicalist Manifesto*, Cambridge: Cambridge University Press.
- Mill, J. S. (1843) *System of Logic*, London: Longmans, Green, Reader & Dyer. Eighth edition, 1872.
- Morgan, C. L. (1923) *Emergent Evolution*, London: Williams & Norgate.

- Papineau, D. (1993) Physicalism, Consciousness, and the Antipathetic Fallacy, *Australasian Journal of Philosophy* 71:2, 169–83.
- Pereboom, D. (2011) *Consciousness and the Prospects for Physicalism*, Oxford University Press.
- Pereboom, D. (This volume) *Consciousness, Physicalism and Absolutely Intrinsic Properties*.
- Papineau, D. (1998) Mind the gap, in J. Tomberlin (Ed.) *Philosophical Perspectives*.
- Papineau, D. (2002) *Thinking about Consciousness*, Clarendon Press: Oxford.
- Papineau, D. 2006. 'Comments on Galen Strawson,' *Journal of Consciousness Studies* 13.
- Perry, J. (2001) *Knowledge, Possibility and Consciousness*, Cambridge, MA: MIT Press.
- Rey, G. (1995) Towards a Projectivist Account of Conscious Experience, in T. Metzinger (Ed.) *Conscious Experience*, Paderborn: Ferdinand Schoningh, 123–42.
- Russell, B. (1927) *The Analysis of Matter*, London: Kegan Paul.
- Sider, T. (2009) 'Ontological realism,' in D. Chalmers, D. Manley & R. Wasserman (Eds.) *Metametaphysics*, Oxford: Oxford University Press, 384–42.
- Sider, T. (2012) *Writing the Book of the World*, Oxford: Oxford University Press.
- Sider, T. (Forthcoming) 'Reply to Schaffer: Metaphysics semantics and multiple realisability,' in 'Symposia on *Writing the Book of the World*,' *Analysis Reviews and Philosophy and Phenomenological Research*.
- Stoljar, D. (2001) Two conceptions of the physical, *Philosophy and Phenomenological Research* 62: 2, 253–81.
- Stoljar, D. (2005) Physicalism and the phenomenal concepts strategy, *Mind and Language*, 20: 5, 469–494.
- Strawson, G. (1994) *Mental Reality*. Cambridge, Mass, London, England: MIT Press.
- Strawson, G. (2003) Real materialism, in L. Anthony & N. Hornstein (Eds.) *Chomsky and his Critics*, Oxford: Blackwell: 48–88.
- Strawson, G. (2006a) Realistic monism: Why physicalism entails panpsychism, in A. Freeman (Ed.) *Consciousness and its Place in Nature*, Exeter: Imprint Academic: 3–31.
- Strawson, G. (2006b). 'Panpsychism? Reply to commentators and a celebration of Descartes,' A. Freeman (Ed.) *Consciousness and its Place in Nature*, Imprint Academic.
- Tye, M. (1995) *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*,  
Cambridge, MA: MIT Press.



Wilson, J. (1999) How superduper does a physicalist supervenience need to be?, *Philosophical Quarterly* 49: 194.

Wilson, J. (2005) Supervenience based formulations of physicalism, *Nous* 39: 3, 426-59.