

Chapter 6 – The Elegant Solution

6.1 Impure Physicalism and Russellian monism

6.1.1 The austerity problem

Pure physicalism is physicalism in conjunction with the view that fundamental reality can in principle be captured in the vocabulary of physical science. In the first half of this book I argued that this view cannot account for consciousness. I want to begin the second half by exploring a problem for pure physicalism which has nothing to do with consciousness, and hence would arise even for an eliminativist about experience.

Pure physicalism fits with a certain popular picture of what physics since the scientific revolution has all been about: a continuous story of progress in understanding the essential 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. There may be a dispute amongst the religious and the secular, but this is generally construed not as a dispute about whether physicists are adequately equipped to capture the material realm, but about whether as well as the material realm studied by physicists there is also an immaterial realm revealed by faith.

However, there is a problem with pure physicalism, even when we set aside the question of whether there is anything 'beyond' the physical realm, and even before we get to worries about whether pure physicalism can account for consciousness. As we discussed in chapters 1 and 2, from Galileo onwards physics has worked with a very austere vocabulary: its predicates express only mathematical and nomic concepts. In chapter 1 I argued that proper respect for science does not require supposing that fundamental reality can be completely captured with such an austere vocabulary. In the first half of this chapter I will argue (independently of considerations pertaining to consciousness) that there *must* be more to fundamental reality than can be captured with physics' austere vocabulary.

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 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, physics provides us with mathematical models that capture the *causal structure* of the universe: reality in so far as it can be captured in a purely mathematico-economic vocabulary. This is very useful information; it enables us to manipulate nature in all sorts of extraordinary ways, leading to extraordinary technology. But mathematical models abstract away from the concrete reality of their subject matter. A mathematical model in economics, for example, abstracts away from the concrete features of real world consumers, such as the nature of their labour and the specific things they buy and sell. Intuitively, wherever there is mathematico-causal structure, there must be some *underlying concrete reality* realising that structure. Physics leaves us completely in the dark about the underlying concrete reality of the physical universe.

On the heterodox 'big picture' I proposed in section 1.1.3, the scientific revolution marks the point when 'natural philosophy' stopped trying to give a complete description of fundamental reality, and instead focused on formulating useful mathematical models of its causal structure. This limited project has been incredibly successful, yielding a technological revolution that has transformed society beyond recognition. Somewhere along the line this impressive record has created in the public mind the conviction that physics is providing us with a complete account of the fundamental nature of the universe. But physics is successful precisely because it is aiming and succeeding at a much less ambitious task.

In the next two sections I will try to demonstrate that this heterodox view is obligatory. I will argue that the language of physics is simply not up to the job of adequately characterising the fundamental nature of matter.

6.1.2 Against causal structuralism

How should philosophers respond to the recognition that physics only tells us about causal structure? One response is the following, 'If physics only gives us causal structure, then causal structure is all there is.' This is the response of the causal structuralist, for whom the complete essential nature of a property is given by its causal profile: by how the property disposes its bearers to affect and to be affected. For the causal structuralist each and every fundamental property is a *pure causal power*. Behaviourists or functionalists believe that the nature of a mental state can be completely captured in causal terms. Causal structuralists generalise this model to the whole of reality, resulting in a kind of *metaphysical behaviourism*. Things are not so much *beings* as *doings*. Pure physicalism is a form of this view.¹

However, there are good reasons to think that a concrete world containing nothing other than pure causal structure is unintelligible. One difficulty that has been pressed by a number of philosophers is that it is hard to see how we can capture the nature of any causal power without referring to properties which are not causal powers.² This is because the nature of a causal power is given with reference to its *manifestation*, i.e. the property it gives rise to. For example, the manifestation of *fragility* is breaking; the manifestation of *flammability* is setting on fire. Arguably the manifestation of a causal power is a *categorical property*, i.e. a property whose nature cannot be captured with only formal or causal predicates. If this is the case, then we cannot fully describe the nature of a power with only formal and causal language. Let us call this the 'specification problem.'

The obvious response for the causal structuralist is to deny that the manifestation of a causal power is a categorical property. Why can't the manifestation of a given causal power be another causal power? The problem is that if the manifestation of causal power F is itself a causal power, call it 'G', then we can only understand the nature of G by understanding *its* manifestation, call it 'H'. If H is also a causal power, then we can understand its nature only by understanding *its* manifestation, call it 'I', and so on *ad infinitum*. Unless at some point we

¹ Bird 2007a, Ellis 2001, 2002, Molnar 2003, Mumford 2004.

² Campbell 1976, Robinson 1982, Blackburn 1980, Armstrong 1997, Heil 2003, Lowe 2006.

find a manifestation that is not itself a causal power, we will never reach an adequate specification of the nature of F.

We can make the specification problem vivid by imagining interrogating a causal structuralist about the nature of some property F:

- 'Right causal structuralist, I want straight answers: What's the essence of F?'
- 'Easy, F is a causal power to give rise to G'
- 'Fair enough, but if the essence of F is given in terms of G, then in order to understand the essence of F I need to understand the essence of G. So what's the essence of G??'
- 'No sweat. G is a causal power to give rise to H.'
- 'Ok...if the essence of G is given in terms of H, then in order to understand the essence of G I need to understand the essence of H. So what's the essence of H???'
- 'H is a causal power to give rise to I...'

Until we reach a specification of a manifestation not given in causal terms, the explanatory buck will continue to be passed and we will never reach an adequate definition of the original causal power.

G. K. Chesterton said that 'We cannot all live by taking in each other's washing.' Bertrand Russell played on this idea in his articulation of the worry outlined above:

There are many possible ways of turning some things hitherto regarded as 'real' into mere laws concerning the other things. Obviously there must be a limit to this process, or else all the things in the world will merely be each other's washing.³

The standard causal structuralist response to the specification problem is to hold that the essence of a causal power is given by its pattern of causal relationships. Consider a world containing three causal powers, F, G and H, such that:

- Fs are disposed to give rise to Gs
- Gs are disposed to give rise to Hs

³ Russell 1927: 325.

- Nothing is disposed to give rise to Fs

We know the nature of F, according to this response, when we know the pattern of causal relationships this property bears to the other two properties co-instantiated with it, i.e. when we know that:

- There is some property that F is disposed to give rise to
- The property F is disposed to give rise to is itself disposed to give rise to some other property
- There is no property that is disposed to give rise to F

According to the causal structuralist all it is for something to be F is for it to bear those causal relationships with two other properties.⁴

It is certainly true that a causal power stands in a pattern of causal relationships to other properties, and perhaps we can uniquely pick out any actual causal power in terms of its pattern of causal relationships. However, to describe such a pattern is not to specify the essence of a causal power. To have a causal power is to be disposed to bring about some change in the world, to make a difference. When I ask what a causal power is I want to know what change it brings about in the world: what property it gives rise to. And I learn what change F brings about in the world when I understand the nature of its manifestation G. But so long as we are restricted to causal predicates, an explanation of G's nature is continuously deferred and never given.

This may sound like it's begging the question. Surely the causal structuralist can just repeat her claim that the essence of a causal power *is* given by the pattern of causal relations. But by definition a causal power is something that gives rise to a certain property, and hence to understand what a causal power is we need to understand the nature of the property it gives rise to. And as we have seen, with only causal or formal predicates the account of that nature can never be given. Describing at a certain level of abstraction the pattern made by the property given rise to, the property it gives rise to, and the property it gives rise to, etc.,

⁴ Hawthorne 2001, Bird 2007b.

does not tell us what the initial causal power *does*; and if you can't tell us what a causal power *does* then you can't tell us what that causal power *is*.

Even setting aside the specification problem, there is a basic intuition that causal powers are too *metaphysically thin* to constitute the complete nature of fundamental concrete objects. A causal power concerns how its bearer points towards other entities and towards its own non-actual but potential manifestation in reaction to those entities. But intuitively fundamental objects must also have a *manifest nature*, a nature that does not consist in such shadowy pointing, but consists in how the object is *in and of itself*. This view is commonly expressed with the analogy that a world of pure powers is like a world in which things are continuously packing their bags for a journey that is never taken: objects continuously change their potentialities, but those potentialities never result in anything *actual*.⁵

This argument relies on the intuition that dispositions are not themselves real enough to constitute the actual nature of objects. Brandishing such brute intuitions may not help to persuade committed causal structuralists. Hugh Mellor wittily compares such metaphysical intuitions to reactionary social attitudes of days gone by: 'Dispositions [i.e. causal powers] are as shameful in many eyes as pregnant spinsters used to be – ideally to be explained away, or entitled by a shotgun wedding to take the name of some decently real categorical property.'⁶ But for those metaphysicians who, after careful reflection, find the notion of a fundamental object of pure potency about as intelligible as the notion of an individual with no properties at all, it is entirely reasonable to look elsewhere for a picture of reality. This is the position I find myself in.

Pure physicalism is a kind of causal structuralism. Hence, if the above considerations persuade us that causal structuralism is incoherent, then we have another reason to reject pure physicalism, a reason that has nothing at all to do with consciousness.⁷

⁵ Armstrong (1997: 80) reports that Professor A. Boyce Gibson made this charge against ordinary language philosophers, and then borrows it to use against causal structuralists.

⁶ Mellor 1974: 157; reported in Bird 2007b.

⁷ The powerful qualities view, i.e. the thesis the causal powers are identical with the categorical properties of objects – Martin 2007 and in Armstrong et al 1996, Martin & Heil 1998, Heil 2003, 2012, Strawson 2008 – is one way of avoiding these difficulties for causal structuralism. I am happy with the view that categorical properties are 'powerful' in the sense that things behave as they do because of their categorical properties,

6.1.3 Introducing impure physicalism

There is an alternative for the physicalist squeamish about causal structuralism, but it means dropping pure physicalism. Instead of taking physical concepts to be transparent concepts revealing the essence of physical properties, we rather take them to be mildly opaque concepts revealing the appearance properties of fundamental physical properties. In other words, we apply the Kripke model of natural kinds like water, outlined in chapter 4, to fundamental physical kinds. Recall that according to Kripke, we pick out water as *the colourless, odourless stuff that fills oceans and lakes and falls from the sky*, whilst water's essential nature as H_2O lies behind this superficial appearance. Applying this model to physical properties: we pick out mass in terms of its dynamical effects – gravitational attraction and resisting acceleration – but perhaps mass has a categorical essential nature lying behind this superficial appearance. By adopting this Kripkean approach, the impure physicalist can avoid the conclusion that mass is a purely causal property, and hence avoid a world in which things are 'merely each other's washing.'⁸

Call this form of physicalism 'impure physicalism'. The impure physicalist gives up on the bold view that completed physics will one day give us the complete metaphysical truth. However, she can still maintain the hope that completed physics will be *referentially adequate*: it could be that all and only physical entities are fundamental entities, even if there is more to the nature of those entities than physics makes known to us.

Does this still count as physicalism? It seems unwise to demand that the physicalist accept such a contentious view as causal structuralism, and most of those who self-describe as physicalists do not. Even if we think that fundamental properties cannot be fully characterised in the vocabulary of physics, one might still take it that they do not involve mentality or value-laden causation, and that they are instantiated by entities at relatively low levels of complexity. In this case, impure physicalism still counts as physicalism according to the definition I developed in chapter 2.

but I think this is better expressed as the view that dispositional properties are grounded in categorical properties. I have difficulty making sense of the idea that the disposition itself is identical with a categorical property, but nothing much will hang on this.

⁸ Lewis (2009) defends impure physicalism and Jackson (1998) expresses sympathy. Arguably Humean physicalism entails impure physicalism.

(It is important to recall at this point that in section 2.1.6 I distinguished between ‘physical’ and ‘material’ facts. Material facts are facts concerning entities at relatively low levels of complexity, and facts grounded in facts concerning entities at relatively low levels of complexity. Physical facts are material facts that don’t involve fundamental mentality/proto-mentality or value-laden causation (I will finally define ‘proto-mentality’ below). We should also keep in mind the distinction I made in the same place between ‘narrowly material/physical’ facts, which are facts that concern entities at low levels of complexity, and broadly material/physical facts, which are facts that are grounded in narrowly material/physical facts).

6.1.4 Introducing Russellian monism

In the nineteen twenties the realisation that there may/must be more to the material world than is revealed by physics suggested to Bertrand Russell and Arthur Eddington a novel solution to the mind-body problem: perhaps the hidden nature of the material world might somehow explain mentality.⁹ The essence of the mind-body problem is that we have two kinds of stuff that don’t seem to fit together, that don’t seem to belong to the same world. Historically dominant solutions have either twisted the mental out of all recognition in order to try to squeeze it into the material, or twisted the material out of all recognition in order to try to squeeze it into an essentially mental world. The Russell/Eddington approach was to put the problem down to our limited understanding of matter, at least in so far as that understanding comes from physical science. We human beings are limited in perception to conceiving of matter in terms of its extrinsic properties, or in terms of what it does; and it is for this reason that we cannot understand matter’s capacity for mentality. When, as it were, God looks down at the material world and conceives of it in terms of its essential nature, there is no ill-fit between matter and mind.

This attractive approach did not fit with the zeitgeist of the latter half of the twentieth century, and does not appear in received histories of twentieth century analytic philosophy of mind.¹⁰ In the twenty first century it is enjoying a rebirth, with discussion focussed on a

⁹ Russell 1927, Eddington 1928.

¹⁰ In these received histories problems with dualism led to behaviourism, problems with behaviourism led to identity theory, problems with identity theory led to functionalism, and problems with functionalism led to dualism. Perhaps it’s time to start over...

Russell inspired view that has become known as ‘Russellian monism.’¹¹ Unlike substance dualism, Russellian monism is a monistic view according to which the explanatorily basic properties of the world are those referred to by fundamental physics.¹² Beyond this we can think of it as having a negative and a positive aspect to its definition:

- *The negative aspect of Russellian monism* – Russellian monists agree with impure physicalists that material properties have a non-structural categorical nature that goes beyond the mathematico-nomic nature captured by physics. For both impure physicalists and Russellian monists, then, we can distinguish between the mathematico-nomic properties in terms of which physics characterises material properties – these are the pure physical properties – and the ‘hidden’ categorical features of material properties that lie behind their pure physical nature – call these the ‘deep nature’ of matter, or ‘deep material properties’.
- *The positive aspect of Russellian monism* – Unlike impure physicalists, Russellian monists think that the deep nature of matter plays a certain special role in the explanation of consciousness. Whilst consciousness *cannot* be explained in terms of pure physical properties, it *can* be explained in terms of the deep nature of matter. (We can make this positive claim more precise by saying that for the Russellian monist: the complete pure physical truth does not a priori entail the complete phenomenal truth, but the complete truth about the deep nature of matter does a priori entail the complete phenomenal truth).

According to the Russellian monist, the reason we find it hard to accept that matter can explain consciousness, the reason it seems that zombies are conceivable, is that we conceive of matter in terms of its extrinsic dispositional nature. If I could magically perceive the deep nature of the fundamental particles making up your brain, it would cease to be conceivable that something with your material nature could lack consciousness: a zombie version of you

¹¹ See, for example, Pereboom 2011, Alter & Nagasawa 2015. There were some forms of the view, although not under that name, defended in the twentieth/early twenty first century, e.g. Feigl 1967, Maxwell 1979, Lockwood 1999, Chalmers 1996, Griffen 1998, Rosenberg 1997, Stoljar 2001, Strawson 2003.

¹² The emergentist Russellian monist – discussed below – thinks that certain macro-level properties and individuals are fundamental. Nonetheless, she still holds that the properties referred to by physics are explanatorily basic, in the sense that they explain the emergence of the fundamental macro-level properties and individuals.

would cease to be conceivable. The essence of the mind-body problem for the Russellian monist is rooted in our scientific ignorance about the deep nature of the material world.

Russellian monism comes in two basic forms. Panpsychist Russellian monists take deep narrowly material facts to be themselves phenomenal facts, i.e. to involve the instantiation of consciousness. On this view, the fundamental constituents of the material world have a consciousness-involving essence, a crude experiential nature which underlies and explains the experiential nature of people and animals.¹³ Panprotopsychist Russellian monists take deep narrowly material facts to involve *proto*-phenomenal properties: properties which are not themselves phenomenal properties, but which are crucial ingredients in facts which *explain* consciousness, in the sense that transparent specifications of such facts a priori entail the phenomenal truths (where that a priori entailment is not wholly in virtue of the pure physical truths thereby entailed). In either case, deep material properties play a crucial role in explaining consciousness.

Given that we are exploring views in which consciousness is to be found in places we don't normally expect to find it, e.g. in fundamental matter, it would be good to have a word for the kind of consciousness we ordinarily believe in, the kind of consciousness we pre-theoretically associate with humans and animals. I call this 'o-consciousness' ('o' for 'ordinary' or 'organism'); o-consciousness is just the kind of consciousness we believe in in everyday life, before we start doing philosophy. Relatedly, I will use the terms 'o-subject' and 'o-phenomenal fact' to refer to conscious subjects and consciousness-involving facts that we have a pre-theoretical commitment to.¹⁴

¹³ I won't be discussing panpsychism in the history of philosophy. See Skrbina 2007 for the panpsychism in the history of Western philosophy.

¹⁴ I take it that Russellian monism is compatible with the powerful qualities view, i.e. the thesis the causal powers are identical with the categorical properties of objects (Martin 2007 and in Armstrong et al 1996, Martin & Heil 1998, Heil 2003, 2012, Strawson 2008). On such a view physics doesn't reveal the deep nature of matter *qua* categorical, and it is the deep nature of matter *qua* categorical that explains o-consciousness. I have some reservations about the powerful qualities view, briefly expressed in footnote 7, but I think the arguments and views of part II of this book could be equally well understood in a powerful qualities framework.

6.1.5 The distinction between Russellian monism and physicalism

Should we count Russellian monism as a form of physicalism? Most Russellian monists agree with physicalists that all and only narrowly material properties are fundamental properties.¹⁵ They think there is more to nature of those properties than can be captured by physics, but then so does the impure physicalist. On the other hand, the Russellian monist thinks that the fundamental properties involve mentality or proto-mentality, which is at odds with the definition of physicalism developed in chapter 2. Some philosophers feel passionately that Russellian monism certainly is a form of physicalism; others are equally passionate that it is not. David Chalmers describes the situation well: ‘whilst the view arguably fits the letter of materialism, it shares the spirit of anti-materialism.’¹⁶

I don’t think there is any fact of the matter here; both ‘physicalism’ and ‘Russellian monism’ are technical terms that we can define as we like to suit our purposes. However, in line with the definition of ‘physicalism’ I developed in chapter 2, I prefer not to think of Russellian monism as a form of physicalism. This way we get a clear(ish) divide between those philosophers who do and those who don’t think we must make distinctive metaphysical commitments in order to account for consciousness. Here are the crucial differences between Russellian monism and type-B and type-A forms of impure physicalism:

The distinction between Russellian monism and type-B impure physicalism

The impure type-B physicalist believes in deep material properties. If she is an identity theorist, she may take deep material properties to be essential constituents of conscious states. But she doesn’t believe that our lack of grasp of deep material properties accounts for the mind-body problem. For the impure type-B physicalist, even if I could magically perceive the deep nature of the material world, I wouldn’t thereby get a transparent explanation of consciousness; even with a God-like grasp of the nature of your brain a zombie version of you would still be conceivable. The impure type-B physicalist is more likely to explain the mind-body problem in terms of the opacity of *phenomenal concepts*: it is

¹⁵ Emergentist Russellian monists do not agree with physicalists on this point.

¹⁶ Chalmers 2002: 265.

because phenomenal concepts don't reveal the physical/functional nature of conscious states that we struggle to see how consciousness could be material.

In contrast the Russellian monist thinks that the deep narrowly material truth a priori entails the o-phenomenal truth, and that it is our lack of grasp of deep material properties that accounts for our difficulty understanding consciousness as material. If human beings were able to perceive the deep nature of matter, it would never have occurred to us that there is a difficulty reconciling matter with mind.

The distinction between Russellian monist and impure type-A physicalism

The type-A physicalist thinks that the complete pure physical truth a priori entails the complete o-phenomenal truth. An *impure* type-A physicalist thinks that the totality of pure physical facts do not exhaust reality: they are merely the appearance properties we use to pick out facts involving deep material properties. However, like all physicalists, type-A physicalists do not think that deep material properties play any special role in explaining consciousness. The complete deep narrow material truth a priori entails the complete o-phenomenal truth, but only in virtue of a priori entailing the complete pure physical truth. In contrast, the Russellian monist thinks that pure physical truths alone do not a priori entail phenomenal truths.¹⁷

We have, then, three different kinds of view, all of which are monistic, and all of which believe in the deep nature of matter:

- *Impure type-A physicalism* – The complete pure physical truth does not exhaustively characterise fundamental reality; nonetheless the complete pure physical truth a priori entails the complete o-phenomenal truth.
- *Russellian monism* – The complete pure physical truth does not a priori entail the complete o-phenomenal truth, but the complete deep narrow material truth does a priori entail the complete o-phenomenal truth.
- *Impure type-B physicalism* – Neither the complete pure physical truth, nor the complete deep material truth, a priori entails the complete o-phenomenal truth;

¹⁷ The definition of Russellian monism offered by Alter & Nagasawa (2015) is problematic because it counts type-A impure physicalists as Russellian monists, as I point out in Goff 2015a.

nonetheless the totality of deep narrowly physical facts grounds the totality of o-phenomenal facts.

6.1.6 The transparency argument against physicalism

Now that we have clearly distinguished between Russellian monism and type-B physicalism, we are in a position to see how the case against physicalism built up in the first half of the book works against type-B physicalism. In chapter 5, I used the following form of the transparency argument to reject pure physicalism and type-A physicalism:

The transparency argument against pure/type-A Physicalism

Premise 1 – If Direct Phenomenal Transparency and either pure physicalism or type-A physicalism are true, then phenomenal concepts reveal their referents to be pure physical states.

Premise 2 – If phenomenal concepts reveal their referents to be pure physical states, then there is no epistemic gap between the pure physical and the experiential.

Premise 3 – There is an epistemic gap between the pure physical and the experiential.

Conclusion 1 – Therefore either Direct Phenomenal Transparency is false, or pure physicalism and type-A physicalism are false.

Premise 3 (Revelation) – Direct Phenomenal Transparency is true.

Conclusion 2 – Pure physicalism and type-A physicalism are false.

The impure type-B physicalist does not accept that fundamental reality is wholly constituted of pure physical facts; rather it is constituted of deep material properties organised into pure physical structures. Furthermore, she may deny that conscious states are pure physical states, taking them to essentially involve deep physical nature. The above argument is therefore ineffective against the impure type-B physicalist.

However, if we assume the Minimal Rationalism I defended in the last chapter, we can show that Direct Phenomenal Transparency is inconsistent with type-B physicalism. Let us remind ourselves of what Minimal Rationalism is:

Minimal Rationalism (2nd formulation) – For any non-fundamental non-singular truth T, a *transparent rendering* of T is a priori entailed by a *transparent rendering* of the fundamental non-singular truths.

A 'transparent rendering' is defined as follows:

A transparent rendering: A transparent rendering of description D is a description which is indiscernible from D except that each term expressing a non-transparent concept is replaced with a term expressing a transparent concept of the same entity.

According to physicalism the fundamental truths are narrowly material truths. Thus we reach the following:

If physicalism is true, then a transparent rendering of the complete (non-singular) narrowly material truth a priori entails a transparent rendering of the complete (non-singular) phenomenal truth.

Remember that for the type-B physicalist (non-singular) narrowly material properties are not phenomenal or proto-phenomenal, and hence it is not the case that transparent renderings of (non-singular) narrowly material truths a priori entail phenomenal truths. Thus, impure type-B physicalism entails the following:

If type-B physicalism is true, then a transparent rendering of the complete (non-singular) deep narrowly material qualitative truth does not a priori entail the complete (non-singular) phenomenal truth.

But Direct Phenomenal Transparency entails that there is no distinction between the complete (non-singular) phenomenal truth (conceived of under direct phenomenal concepts) and a transparent rendering of the complete (non-singular) phenomenal truth.

We can therefore offer the following Transparency Argument against type-B physicalism (taking the phenomenal truths to be conceived of under direct phenomenal concepts):

The transparency argument against type-B physicalism

Premise 1 – If physicalism is true, then a transparent rendering of the complete deep (non-singular) narrowly material truth a priori entails a transparent rendering of the complete (non-singular) phenomenal truth (from Minimal Rationalism and the definition of physicalism).

Premise 2 – If type-B physicalism is true, then a transparent rendering of the complete (non-singular) deep narrowly material truth does not a priori entail the complete (non-singular) phenomenal truth.

Premise 3 (Direct Phenomenal Transparency) – The complete (non-singular) phenomenal truth is identical with a transparent rendering of the complete (non-singular) phenomenal truth.

Conclusion 1 – Therefore, if type-B physicalism is true, then (the complete (non-singular) deep narrowly material truth a priori entails the complete (non-singular) phenomenal truth) & ~ (the complete (non-singular) deep narrowly material truth a priori entails the complete (non-singular) phenomenal truth).

Conclusion 2 – Type-B physicalism is false.

Now that we have refuted physicalism, I turn to exploring Russellian monism as an alternative way of finding a place for consciousness.

6.3 Varieties of Russellian Monism

6.3.1 Constitutive and emergentist forms of Russellian monism

Russellian monists think that the deep nature of the matter explains consciousness: that the deep material truth a priori entails the o-phenomenal truth. In line with the Minimal Rationalism I defended in chapter 5, I am inclined to think that if fact X constitutively grounds fact Y then a transparently rendered proposition expressing X a priori entails a transparently rendered proposition expressing Y.¹⁸ However, I don't believe this conditional runs in the other direction; that is to say, I don't believe that if a transparently rendered proposition expressing X a priori entails a transparently rendered proposition expressing Y, then X constitutively grounds Y.

¹⁸ This claim is consonant with, although not quite entailed by my principle of Minimal Rationalism.

Some examples should serve to make the point. We can move a priori from *God's willing that there be light* to *there being light*, and furthermore (in the imagined scenario in which both obtain) the latter state of affairs obtains *because* the former obtains; in this sense the state of affairs of there being light is intelligibly produced by the state of affairs of God willing that there be light. Nonetheless, there being light is not *constitutively grounded* in God's willing that there be light. Recall that by definition if X constitutively grounds Y, then Y is nothing over and above X. But in willing that there be light God creates new being; and hence the fact that there is light is something over and above – and hence not constitutively grounded in – the fact that God wills that there be light. To take a non-theological example: causal structuralism in conjunction with determinism entails that we can move a priori from facts about the past to facts about the future (if causal structuralism is true, then knowing which properties an individual instantiates entails knowing its causal dispositions); and yet clearly facts about the future are something over and above, and hence not constitutively grounded in, facts about the past.

We can distinguish, therefore, between two forms of intelligible production: *constitutive grounding*, in which the produced is nothing over and above the producer, and *intelligible causation*, in which the produced is something over and above the producer. The fact that there are 11 drunk teenagers revelling in a room intelligibly produces the fact that there is a party; the product is nothing over and above the producer and we have a case of constitutive grounding. Assuming causal structuralism and determinism, the fact that a lump of sugar is dropped in a cup of tea intelligibly produces the fact that there is sugary tea; the product is something over and above the producer and we have a case of intelligible causation.

The case of intelligible causation considered above is a case of diachronic causation, i.e. the causation happens across time. However, there could also be synchronic intelligible causation, for example, in which a fundamental lower-level property intelligibly causally sustains a fundamental higher-level property. This would be a case of *intelligible emergence*: fundamental lower-level properties bringing into being fundamental higher-level properties,

in such a way that one could in principle move a priori from the instantiation of the lower-level property to the instantiation of the higher-level property.¹⁹

In the light of this, we can distinguish two forms of Russellian monism. The *constitutive Russellian monist* believes that the deep narrowly material facts *constitutively ground* the o-phenomenal facts, and hence that the o-phenomenal facts are nothing over and above the deep narrowly material facts. The *emergentist Russellian monist* believes that deep narrowly material facts *intelligibly cause* the o-phenomenal facts, and hence that the o-phenomenal facts are something over and above the deep narrowly material facts.²⁰

Perhaps the most natural way of thinking about emergentist Russellian monism is as a kind of layered view of nature involving upwards synchronic causation; the kind of intelligible emergence discussed above. The micro-level subjects (or proto-subjects, i.e. bearers of proto-phenomenal properties) in the brain causally bring into being the o-conscious subject; micro-level subjects and o-subject then co-exist, with the former causally sustaining the existence of the latter. At death, or perhaps during dreamless sleep, the micro-level subjects cease to be arranged in an o-subject sustaining way, and the o-subject consequently ceases to exist. Call this the ‘layered’ form of emergentist Russellian monism.²¹

Traditionally emergentism has been thought of as involving brute rather than intelligible causation. However, assuming the coherence of intelligible forms of causation opens up the possibility of intelligible emergentism. Indeed, given its definitional commitment to an a priori entailment from the deep material truths to the o-phenomenal truths, intelligible emergentism is the only form of emergentism consistent with Russellian monism.

There is an alternative to the layered conception of emergentism, defended by William Seager and Hedda Hassel Mørch.²² On this view, an o-subject comes to be as the result of micro-(proto)subjects *fusing* together to form a single entity, giving up their individual

¹⁹ It is tempting to analyse the distinction between grounding and causation in terms of whether or not the product is something over and above the producer. However, this would require denying that there could be cases of *non-constitutive* grounding; the ‘Moorean emergentism’ considered in chapter 1 would turn out to be a case of intelligible causation (we would perhaps have to deny the coherence of the Moorean metaethical view). I am not averse to this possibility, but haven’t thought it through enough to defend it here.

²⁰ This distinction between ‘constitutive Russellian monism’ and ‘emergentist Russellian monism’ comes from Chalmers 2015.

²¹ Rosenberg 2004 and Brüntrup 2016 defend layered emergentism.

²² Seager Forthcoming, Mørch 2014.

existences in the process. Micro and o-(proto)subjects do not co-exist: the o-subject – identical with the brain of the conscious organism – is a perfectly unified entity, lacking fundamental parts of any kind.²³

How could it be plausibly held that a human brain lacks fundamental parts? The brain has rich structure, it has different properties associated with what we think of as its different regions; it's not just a single homogenous blob. How can any of this be made consistent with the thesis that the brain is a mereologically simple entity? One way to make sense of this is with reference to Josh Parsons's work on *distributional properties*. Distributional properties are properties which concern how a region of space is 'filled in'; examples of distributional properties are *being polka dotted*, or *being striped*.²⁴

It is natural to think of distributional properties of an object as grounded in the properties of its parts, e.g. the carpet has the distributional property of *being polka dotted* because of the colours of its various bits. However, Parsons suggests that it is at least coherent to suppose that fundamental distributional properties might be had by objects which lack parts. There could have been fundamental particles with the brute property of *having a polka dotted distribution of mass*, despite having no parts to ground that property.²⁵

If this all makes sense, then it is coherent to suppose that post-fusion a mereologically simple object is formed, which nonetheless has rich structure in virtue of instantiating complex distributional properties. What we think of as different parts of the post-fusion brain are actually different aspects of the distributional properties instantiated by the fundamental brain. One could of course remove a bit of someone's brain, but only by causing the fused whole to dis-unify into many distinct entities, some of which are then removed. Call this view 'non-layered emergentist Russellian monism', or 'fusionism' for short.

Even if coherent sense can be made of the brain's lacking parts, one might nonetheless object to the view on the grounds that it is wildly counter to common sense. But the

²³ In fact Seager and Hassel propose panpsychist forms of this view, but we can easily imagine a panprotopsychist version.

²⁴ Parsons 2004.

²⁵ I myself have doubts that extended simples are coherent. But I certainly think that the notion of an extended fundamental particle with dependent parts is coherent. We shall return to these issues in chapter 9.

fusionist could allow that the brain has *non-fundamental* parts, grounded in the distributional properties of the fundamental whole. Perhaps facts about the whole brain and its distributional properties ground facts about the brain's parts. It is quite common to suppose that macroscopic parts of the brain are non-fundamental, grounded in facts about micro-level parts; the fusionist just offers an alternative ground for these non-fundamental parts. And whilst common sense may insist that micro-level parts of the brain exist, it does not have firm feelings on their ontological status.²⁶

In contrast to both forms of emergentism, the constitutive Russellian monist denies the fundamentality of o-phenomenal facts. The constitutive Russellian monist picture of the universe is structurally very similar to that of the physicalist, in that both take fundamental reality to be wholly constituted of facts concerning individuals and properties at low levels of complexity.²⁷ The constitutive Russellian monist differs from the physicalist only in her insistence that low-level entities must be conscious or proto-conscious in order to account for the existence of o-consciousness.

6.3.2 The causal exclusion problem

The structural similarity with physicalism allows the constitutive Russellian monism to avoid a much discussed difficulty, associated with Jaegwan Kim, known as the causal exclusion problem.²⁸ This problem arises from the difficulty reconciling two theses many philosophers find attractive:

Micro-level causal closure – Any caused event has a sufficient micro-level cause. This is roughly equivalent to thesis that completed physics will in principle be able to causally explain any caused event.²⁹

²⁶ It is more natural to take this form of grounding as *grounding by subsumption*, which we will discuss in chapter 9, rather than *grounding by analysis*, which we discussed in chapter 2.

²⁷ Unless they adopt cosmopsychism; the view we will explore in chapter 9. The cosmopsychist's picture of the world is structurally similar to that of the priority monist physicalist.

²⁸ Kim 1989, 1993a, 1998, 2005. An early version of the point is due to Malcolm 1968.

²⁹ Two qualifications: (A) Causal explanation may be non-deterministic, (B) Micro-level causal closure does not ensure that micro-level events *directly* cause macro-phenomenal events; more plausibly if micro-level event MICRO causes macro-level event MACRO, then it does so indirectly by causing the micro-level event in which MACRO is grounded. Some might prefer to characterise the principle in terms of 'causal explanation' rather than 'causation'.

Causal efficacy of o-consciousness – O-phenomenal facts have causal effects in the material world, e.g. my pain causes my screaming and running away.

Assume the truth of both of these theses. Now consider some material event E caused by an o-phenomenal fact, e.g. my screaming and running away caused by my feeling pain. Given causal closure, E must have a sufficient micro-level cause M. And yet E is also caused by the fact that I am in pain. Hence, E seems to have too many causes: M and the fact that I am in pain.

On the face of it, accepting both of the above theses seems to lead to *systematic over-determination*: every single event caused by an o-phenomenal fact has two sufficient causes. Whilst there may be fluky one-off situations in which events are over-determined – Sarah's death is over-determined when she happens to be shot and struck by lightning at the same time – it would be deeply theoretically unattractive to suppose that every single event that results from o-consciousness is over-determined in this way. However, many philosophers are reluctant to deny either of the above theses, which in conjunction seem to push us in this direction.

The physicalist arguably avoids this worry by holding that o-conscious facts are constitutively grounded in micro-level facts. Consider the following case of an event with two causes:

- My sleepless night was caused by the fact that there is a party upstairs
- My sleepless night was caused by the fact that there are people dancing and drinking upstairs.

In some sense my sleepless night has two causes:

- The fact that there is a party upstairs.
- The fact that people are dancing and drinking upstairs.

However, this is not a worrying case of overdetermination as the fact that there is party upstairs is nothing over and above the fact that there are people dancing and drinking upstairs.

One more example:

- The baseball smashed the window.
- Particles arranged baseball-wise acting in concert smashed the window.

In some sense the smashing of the window has two causes:

- The fact that the baseball impacted on it.
- The fact that particles arranged baseball-wise impacted on it.

But again, this is not a worrying case of overdetermination so long as the fact that there is a baseball impacted on the window is nothing over and above the fact that the particles arranged baseball-wise impacted on the window.

Call such cases of over-determination, in which one of the causes is nothing over and above the other, cases of 'superficial over-determination'. The case of the woman's death being caused by both a bullet and a bolt of lightning is not a case of superficial over-determination, as the two causes are grounded in distinct facts, neither of which is grounded the other. Call such cases 'deep over-determination.'

The physicalist avoids the causal exclusion problem as she thinks that the cases of over-determination of behaviour implied by micro-level casual closure are really cases of superficial over-determination. Return to the case we started with:

- The fact that I am in pain caused me to scream and run away.
- Micro-level fact M caused me to scream and run away.

The physicalist is likely to hold that my pain is grounded in M. Perhaps M is a micro-level event underlying the neurophysiological state with which my pain is identical. Or perhaps my pain is a functional state realised by a neurophysiological state which is in turn grounded in M. In either case, my pain is nothing over and above the fundamental physical goings on that cause my behaviour, and hence worries about over-determination evaporate.

The constitutive Russellian monist, given the structural similarities between her view and physicalism, is able to give the same solution. The only difference would be that, as the constitutive Russellian monist tells the story, M is a micro-level event involving consciousness or proto-consciousness:

- The fact that I am in pain caused me to scream and run away.
- M (= the fact that certain micro-level entities with certain (proto)phenomenal states are related in such and such a way) caused my screaming and running away.

According to a standard form of constitutive Russellian monist the fact that I am in pain is grounded in, is nothing over and above, some fact concerning certain of my fundamental material parts and their (proto)phenomenal properties. If we identify that fundamental fact with the micro-level cause of my pain behaviour, then the over-determination of my pain behaviour turns out to be superficial and unproblematic.³⁰

This advantage is not shared by layered emergentist Russellian monism, as on this view o-phenomenal facts are something over and above the micro-level (proto)phenomenal facts which produce them; the relationship between the micro-level (proto)phenomenal and the o-phenomenal is intelligible causation rather than grounding. The o-phenomenal fact of my feeling pain co-exists with the (proto)consciousness-involving micro-level fact that sustains it, each being a fundamental fact in its own right. If the two facts are both sufficient causes of my screaming and running away, then this is clearly an instance of deep, and so problematic, over-determination (assuming micro-level causal closure).³¹

The situation is slightly more complicated for the fusionist. There is no over-determination, as post-fusion there are no longer any fundamental micro-level events taking place in the conscious brain; there is just the fundamental macro-level brain. Hence, unlike layered emergentism there aren't two fundamental levels to potentially deeply over-determine behaviour. But fusionism appears to be inconsistent with micro-level causal closure in a more direct way: if there is no micro-level cause of my screaming and running away (as the fundamental brain caused the action) then micro-level causal closure is false.

Fusionist Hedda Hassel Mørch has explored ways of getting around this.³² As we noted earlier, the fusionist may hold that the brain has *non-fundamental parts*, grounded in the fundamental brain and its distributional properties. It is thus an option to hold that there is a

³⁰ A similar argument is made in Chalmers 2015.

³¹ Worries about over-determination would also be removed if the two causes were on a single causal chain, but this does not seem to be the case here.

³² Mørch 2014.

sufficient cause of my screaming and running away among these non-fundamental micro-level goings on. We would then get a case of superficial over-determination, this time because the micro-level cause is grounded in the macro-level cause rather than vice-versa.

Mørch concedes, however, that this response is not straightforward. Although not breaking the letter of micro-level causal closure, fusionism seems inconsistent with its spirit. The proponent of micro-level causal closure doesn't just think that for every event there happens to be some sufficient cause at the micro-level. The conviction in the background is that facts at the micro-level are 'running the show', i.e. are the ultimate causal source for the evolution of the universe. We will examine in chapter 9 whether there is any justification for this conviction, but for now we can note that fusionism is inconsistent with it: post-fusion it is the macro-level brain rather than anything at the micro-level that is running the show. Pre-fusion micro-level parts are running the show; post-fusion the macro-level brain is running the show.

We should further note that the post-fusion micro-level parts are not numerically identical with the pre-fusion micro-level parts, as the former – being grounded in facts about the newly created fundamental macro-level brain – only come into existence at the moment of fusion. The fusionist might insist that the post-fusion micro-level parts behave according to exactly the same laws as did the pre-fusion micro-level parts; in this way they would preserve the letter but not the spirit of micro-level causal closure. However, it is deeply strange that these new post-fusion entities should continue to behave exactly as the numerically distinct pre-fusion micro-level entities behaved, as if no change has taken place.

What takes place when an o-conscious entity comes into existence, according to the fusionist, is a radical change in nature. A huge number of micro-level entities pass out of existence and are replaced by a fundamental macro-level entity; we go from a situation in which trillions of things at the micro-level are in the driving seat, to a situation in which a single macro-level entity is in the driving seat. It would be weird if that change in driver didn't show up empirically; if the brand new macro-level entity continues to make the micro-

level run just as it did when it ran itself. We would be left with the sense that nature was conspiring to hide this radical change from us.³³

There is strong pressure, then, for the emergentist, of either the layered or non-layered variety, to deny either *causal efficacy of o-consciousness* or *causal closure of the micro-level*. The win-win solution to the causal exclusion problem that is available to the constitutive Russellian monist is not available to the emergentist Russellian monist. This is widely seen to be a significant *prima facie* advantage to constitutive versions of Russellian monism. How great an advantage this really is depends on how great the evidence is that the micro-level is causally closed. This will be explored in chapter 9.

We can see, then, that constitutive Russellian monism is a very attractive position.

Physicalism pushes us to choose from a set of *prima facie* unattractive options:

- *Phenomenal opacity* – The thesis that phenomenal concepts are opaque.
- *Analytic functionalism* – All mental concepts admit of causal analysis, which seems to entail that phenomenal concepts don't exist.
- *Dual Carving* – There are two conceptually distinct ways of understanding the essences of certain fundamental material properties.

Dualism or emergentism also has its own *prima facie* unattractive options:

- *Epiphenomenalism* – O-consciousness has no causal efficacy.
- Denial of micro-level causal closure (At least many philosophers think of this option as unattractive; we will discuss in chapter 9 whether it really is).

Constitutive Russellian monism elegantly avoids all of this pain. The necessitation of the o-phenomenal by the micro-level is perfectly intelligible; it is merely the opaque grasp of matter afforded by the physical sciences that bars us from appreciating this. And because o-phenomenology is constitutively grounded in micro-level facts, we can happily accept its causal efficacy without denying micro-level causal closure. The constitutive Russellian monist can have her cake and eat it too.

³³ The essence of the concern I am raising here for fusionism is from Mørch 2014.

6.3.2 Panqualityism

There is a form of Russellian monism – dubbed ‘panqualityism’ by Herbert Feigl and prominently defended in recent times by Sam Coleman – that doesn’t easily fit into the categories discussed above.³⁴ If fusionism comes pretty close to substance dualism (like Descartes the fusionist thinks the mind is fundamental and indivisible – it’s just that the brain is too!), then panqualityism comes pretty close to physicalism. In fact, panqualityism is perhaps best seen as a kind of hybrid of Russellian monism and type-A physicalism. The type-A physicalist explains consciousness in terms of the pure physical properties of matter. The Russellian monist explains consciousness in terms of the deep nature of matter. The panqualityist analyses consciousness into two different aspects, one of which is explained by the pure physical properties of matter, the other of which is explained by the deep nature of matter.

The two aspects the panqualityist analyses consciousness into are *quality* and *subjectivity*. Quality is the quality-involving nature of consciousness. Intuitively, consciousness involves qualities: the redness involved in our experience of red, the sweet smelliness involved in an experience of lavender, the spiciness of the taste of paprika. As discussed in chapter 1, the scientific revolution arguably begins with Galileo taking these qualities out of the material world. However, Galileo didn’t eliminate these qualities – he put them inside the mind. Similarly, the panqualityist doesn’t think we can get rid of these qualities as their reality is evident in our experience; she follows Galileo in putting them inside the mind, which she takes to be identical with the brain. The redness in your experience, the spiciness in your conscious experience of paprika, the sweet smell in your experience of lavender; for the panqualityist all of these qualities characterise the matter of your brain. Matter is as it were multi-coloured. It is precisely because a pure physical description of your brain doesn’t capture these qualities that pure physicalism cannot be true.

Subjectivity, the second aspect of consciousness, is a matter of these qualities being *for the subject*. When I’m looking at something red, the redness is *for me* in the sense that it

³⁴Feigl (1960) credits it to a conversation with Stephen C. Pepper. Versions of the view were held by William James (1904), Ernst Mach (1886), Bertrand Russell (1921) and Peter Unger (1999). Coleman elaborates and defends the view in his 2012, 2013, 2015 and 2016.

characterises my experience: it's part of what it's like to be me.³⁵ It is this second aspect of consciousness that the panqualityist is effectively a type-A physicalist about. Pure physical structure alone can't account for qualities, but once we have qualities organised in the right kind of pure physical structures, we thereby get subjectivity: we get qualities that are *for a subject*. Facts involving qualities arranged in pure physical structures ground facts about subjective experience. The panqualityist may, for example, define consciousness as *awareness of qualities*, where the relevant notion of awareness admits of functionalist analysis. For me to have a conscious experience of red is for me to be aware of a red quality in my brain; for me to be aware of a red quality in my brain is for that red quality to stand in the right kind of causal arrangement with other bits of my brain.

The resulting view is not type-A physicalism, as the pure physical truths alone cannot account for consciousness: you need the qualities. Nor is it type-B physicalism, as the fundamental material truths – qualities and pure physical structures – a priori entail the phenomenal truths. But it's not straightforwardly a form of panprotopsychism either. For panprotopsychists the mystery of consciousness is rooted in our ignorance of the deep nature of matter. For panqualityists this is half right: the qualitative aspect of consciousness is mysterious because of our ignorance of the qualitative aspect of matter, but the subjective aspect of consciousness can be explained away in standard physicalistic terms. In order to distinguish panqualityism from the undiluted physicalist views we have already rejected, I shall think of panqualityism as a form of constitutive Russellian monism – for one thing it is arguably the view of Russell himself!³⁶ – but one distinct from either panpsychism or panprotopsychism.³⁷

There is a sense in which panqualityism is closer to panpsychism than panprotopsychism: for both panpsychists and panqualityists the deep nature of matter consists of qualities akin to those we find in our experience. Presumably the qualities we find in o-experience are not

³⁵ Something like this distinction between subjectivity and quality is made in Kriegel 2008 (pp. 45-57), but not in the service of a defence of panqualityism.

³⁶ Russell 1927. There is controversy about how exactly to interpret Russell's view and its relationship to contemporary Russellian monism; see the essays of Wishon and Stubenberg in Alter & Nagasawa 2015.

³⁷ Tom McClelland (2012) defends a not unrelated view. He also distinguishes between quality and subjectivity, and proposes a type-A physicalist-style account of the former and a Russellian-style account of the latter. However, McClelland favours a kind of *noumenalism* (see next chapter) about the deep nature of matter.

precisely the same as those instantiated in fundamental matter, but they are of the same general type. The difference is that for the panqualityist the qualities such we find in experience can exist unexperienced, and they do so exist in fundamental matter. Before the highly developed point in evolution when qualities come to be arranged in the right functional structures, they existed unexperienced.

This can be seen to lead to conceptual difficulties for the view. Can we make sense of the qualities in our experiences existing unexperienced? Consider a pain in your leg. Could that very quality exist unfelt? Arguably, the qualities in our experience just are, in their essential nature, experience-characterising properties. The quality I find when I attend to the pain in my leg is *what it's like to feel pain (of a certain kind)*. How could *what it's like to feel pain* be instantiated without anyone feeling pain?

Moreover, even if we can make sense of the qualities in our experience existing unexperienced, it's not clear that merely putting them into certain causal structures will render them experienced. As we have seen, opposition to physicalism is often motivated by the conceivability of the pure physical truth obtaining in the absence of consciousness. Panqualityists add fundamental qualities into the mix to try to account for consciousness. But the problems is that, even after we have added fundamental qualities, the absence of consciousness is arguably still conceivable. It seems that we can conceive of a 'panqualityist zombie world': a world just like our world in its pure physical structures, composed of richly qualified matter, but in which there is no subjective experience whatsoever (i.e. none of the qualities are *for* a subject). If such a world is conceivable, panqualityism shares the physicalist's inability to account for consciousness.³⁸

For both of these reasons, panqualityism seems to me an unhappy halfway house between physicalism on the one hand and panpsychism/panprotopsychism on the other. I think the view deserves further consideration: if the challenges above could be responded to, then it offers an attractive theory of reality, avoiding some of the worries with panpsychism (the subject-summing problem) and panprotopsychism (the threat of noumenalism) which we

³⁸ Chalmers (2015) makes this argument.

will consider in the next chapter. However, the intuitions which lead one to reject physicalist accounts of quality are unlikely to rest content with physicalist accounts of subjectivity:

I am in blood stepped in so far that should I wade no more,
Returning were as tedious as go o'er.³⁹

With some reservations, therefore, I will set aside panqualityism, and focus only on panpsychism and panprotopsychism in what follows.

6.3.3 A promising view

There are two big divisions amongst Russellian monists:

- Panpsychist forms versus panprotopsychist forms
- Constitutive forms versus emergentist forms

These divisions are orthogonal, leading to four forms of Russellian monism. By dividing emergentist forms into layered and fusionist varieties, and adding panqualityism as a distinct form of constitutive Russellian monism, we end up with seven main forms of Russellian monism:

The seven forms of Russellian monism

- Constitutive panpsychism
- Layered emergentist panpsychism
- Fusionist panpsychism
- Constitutive panprotopsychism
- Layered emergentist panprotopsychism
- Fusionism panprotopsychism⁴⁰
- Panqualityism

Physicalist forms of monism cannot account for consciousness. The hope is that Russellian monism may fare better in this regard. Moreover, Russellian monism provides a prima facie more unified and elegant picture of nature than other non-physicalist options:

³⁹ Macbeth: Act III, Scene IV.

⁴⁰ As far as I know this view has not thus far been defended by anyone.

- *Substance dualism* – Fundamental reality is divided into two very different kinds of thing: utterly non-mental matter on the one hand, and utterly immaterial o-subjects on the other.
- *Brute emergentism* – O-consciousness arises from non-conscious matter in virtue of brute laws of nature.
- *Idealism* – Fundamental reality is exclusively composed of immaterial o-phenomenal subjects. Inanimate matter is grounded in facts involving the experiences, and dispositions to have experiences, of fundamental o-phenomenal subjects.⁴¹

Russellian monism provides a much more unified picture of reality than substance dualism. In contrast to brute emergentism, Russellian monism takes there to be intelligible connections from narrowly material facts to the o-phenomenal facts, thus giving us a more elegant picture of the world.

Idealism is at least as economical as Russellian monism. The problem with idealism is that it is hard to see how it can account for the structure of o-experience, and the commonality of experience between different o-phenomenal minds. Why is it that you and I both experience a common world? Why is it that we both see a tree in the quad? The most straightforward explanation is that there is a tree in the quad, the existence of which is not dependent on o-phenomenal minds, and that that single tree causes both of our experiences; it is the common o-mind-independent world which explains the commonalities in experience between different o-minds. The Russellian monist, like the physicalist, accepts this straightforward explanation (for the panpsychist Russellian monist, the tree is dependent on certain fundamental narrowly material subjects, but it is not dependent on any o-subjects). The idealist in contrast seems to be left with a series of seemingly disconnected facts about what each individual subject is disposed to experience, with no explanation of how those facts ‘hang together.’

Thus, once we have rejected physicalism, Russellian monism seems to be the most promising metaphysical theory, at least for those committed to respecting the Consciousness Constraint. Even its opponents have expressed admiration for its virtues; physicalist Alyssa Ney says of it ‘This proposal strikes me, suspending disbelief about

⁴¹ For a good recent defence of idealism see Robinson 2016. See also Foster 1982.

the...theses that lead up to it, as at least as bold and exciting as Newton's proposed identification of terrestrial and cosmic reality.'⁴²

Constitutive Russellian monism is perhaps the ideal, given its capacity to reconcile the causal efficacy of human and animal feelings and emotions with micro-level causal closure.

However, both emergentist and constitutive views provide us with an elegant and unified picture of the world, in which all entities intelligibly arise in one way or another from the interactions of micro-level entities. Russellian monism is our best hope for finding a place in nature for consciousness.

In this next two chapters, we will explore some major challenges for the view.

⁴² Ney 2015: 349.