

Durham E-Theses

Rethinking the Motivation for Russellian Panpsychism: An Investigation into Russellian Monism, Conceivability Arguments, and the Evolution of Consciousness.

LEE, PHILLIP,DUNCAN

How to cite:

LEE, PHILLIP,DUNCAN (2025) *Rethinking the Motivation for Russellian Panpsychism: An Investigation into Russellian Monism, Conceivability Arguments, and the Evolution of Consciousness.* , Durham theses, Durham University. Available at Durham E-Theses Online:
<http://etheses.dur.ac.uk/16151/>

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

RETHINKING THE MOTIVATION FOR RUSSELLIAN PANPSYCHISM

Duncan Lee

Abstract:

This thesis proposes that Russellian panpsychism provides a better solution to the Mind-Body Problem than the currently dominant approach: physicalism. While dualism and physicalism have been the most traditionally popular forms of response, Russellian panpsychism is proposed to harness the virtues of both, whilst avoiding their vices (Chalmers, 2016a). Unfortunately, Russellian panpsychism suffers its own problem – the Combination Problem – which is often understood to diminish its appeal (James, 1890; Seager, 1995; Chalmers, 1996; Goff, 2006; Rosenberg, 2004; Coleman, 2016). I shall propose that even with the Combination Problem looming, Russellian panpsychism can still offer the best available solution to the Mind-Body Problem.

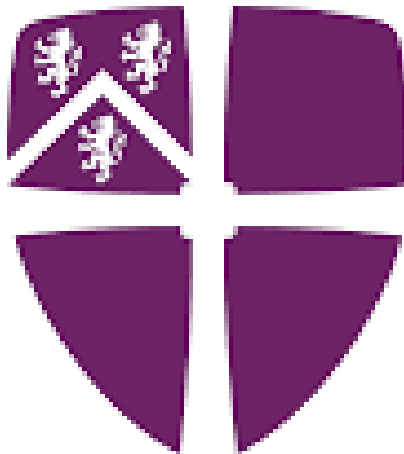
While the ‘Conceivability Argument’ against physicalism is typically used to help motivate this conclusion, I shall argue that Russellian panpsychism is vulnerable to its own version of the argument, and that Conceivability Arguments are fundamentally flawed in-and-of-themselves (Goff, 2009; Cutter, 2019; Chalmers, 1996; 2010; Goff, 2017). With this in mind, and after recommending that they are relinquished from the debate, I shall structure a new argument for Russellian panpsychism and against physicalism, concerning the evolution of consciousness. This argument will serve two functions: (1), it shall serve as a safeguard to Russellian panpsychism, such that even with the Combination Problem looming, Russellian panpsychism can still provide a more attractive theory of consciousness than physicalism; (2), future responses to the Combination Problem can build from the logic of my argument, such that this thesis can serve as a foundation for later authors to tackle the problem effectively.

RETHINKING THE MOTIVATION FOR RUSSELLIAN PANPSYCHISM

AN INVESTIGATION INTO RUSSELLIAN MONISM, CONCEIVABILITY
ARGUMENTS, AND THE EVOLUTION OF CONSCIOUSNESS

Thesis submitted in accordance with the requirements of Durham University for the degree of
Doctor of Philosophy

Duncan Lee



Philosophy Department

Durham University

March 2025

Table of Contents

Statement of Copyright	iii
Acknowledgements	iv
Introduction	1
Chapter 1	
The Mind Body Problem: The Traditional Responses ~ Physicalism and Dualism	7
1.1 The Mind-Body Problem	8
1.2 Dualism	13
1.3 Physicalism.....	21
1.4 Final Remarks	31
Chapter 2	
Russellian Monism: A Potential Midground Solution	34
2.1 The Traditional Formulation	35
2.1.1 The Negative Component	36
2.1.2 The Positive Component	42
2.1.3 The Midground Between our Traditional Theses	44
2.2 The Variants of Russellian Monism	47
2.2.1 Russellian Panpsychism versus Panprotopsychism	48
2.2.2 Constitutive versus Emergent Russellian monism	53
2.2.3 Micropsychism versus Cosmopsychism	60
2.3 Considerations and Concerns	64
2.3.1 Is Russellian Panpsychism at Odds with Science?	65
2.3.2 The Combination Problem	69
2.4 Concluding Remarks	77
Chapter 3	
The Conceivability Argument: Against Physicalism, and for Russellian Panpsychism	80
3.1 The Conceivability Argument Against Physicalism	81
3.1.1 The Simple Conceivability Argument	81
3.1.2 The Complex Conceivability Argument	89
3.2 Defending Physicalism	109
3.3 Russellian Panpsychism's Alleged Immunity	117
3.4 Concluding Remarks	120
Chapter 4	
The Misuse of Conceivability Arguments: Applications and Contradictions	122
4.1 Conceivability Arguments against Russellian Panpsychism	124
4.2 Defending Russellian Panpsychism	132
4.3 The Universal Applicability of Conceivability Arguments	137

4.4 The Issue with Conceivability Arguments	153
4.4.1 Counterfactual Conceivability Arguments	154
4.4.2 Addressing the Issue	174

Chapter 5

An Evolutionary Argument Against Physicalism: Physicalism's Reduction into Panpsychism..	186
--	-----

5.1 The Problems that Remain	187
5.2 The Traditional Argument from Evolution	190
5.3 A New Argument from Evolution Against Physicalism	192
5.3.1 The Precision of Conscious Experience	194
5.3.2 Evolution, Saltation, and Biopsychism	216
5.3.3 The Issue with Biopsychist Physicalism	226
5.4 Further Considerations	240
5.4.1 Prospective Responses to the Combination Problem	241
5.4.2 Further Contextualisation	253
5.5 Concluding Remarks	265

Conclusion	271
-------------------------	-----

Bibliography	277
---------------------------	-----

Statement of Copyright

The copyright of this thesis (including any appendices or supplementary materials to this thesis) rests with the author, unless otherwise stated.

© Duncan Lee (2025).

This copy has been provided under licence to the University to share in accordance with the University's Open Access Policy, and is done so under the following terms:

- this copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- any quotation from the work (e.g. for the purpose of citation, criticism or review) should be insubstantial, should not harm the rights owner's interests, and must be accompanied with an acknowledgement of the author, the work and the awarding institution.
- the content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

Acknowledgements

First, I would like to thank my fiancé, Cat Gillen, for her unwavering support throughout the research and writing of this thesis. The countless dog walks spent discussing philosophy have been instrumental to my writing and are some of my most cherished memories. Second, I would like to thank my supervisor, Philip Goff, for his sustained help, encouragement, and friendship. I would also like to thank Peter West and Jack Symes for their support and invaluable philosophical input, the ‘Wednesday Club’ for their contribution to our thriving PhD community, and the staff at Head of Steam for making sure our glass is always more than half-full. Finally, I would like to thank my parents for their unconditional support, alongside the encouragement that I always follow my passions.

This work was funded by AHRC Northern Bridge Consortium [AH/RO12415/1]

Introduction

This thesis proposes that Russellian panpsychism provides a better solution to the Mind-Body Problem than the currently dominant approach: physicalism. While dualism and physicalism have been the most traditionally popular forms of response, Russellian panpsychism is proposed to harness the virtues of both, whilst avoiding their vices (Chalmers, 2016a). Unfortunately, Russellian panpsychism suffers its own problem – the Combination Problem – which is often understood to diminish its appeal (James, 1890; Seager, 1995; Chalmers, 1996; Goff, 2006; Rosenberg, 2004; Coleman, 2016). I shall propose that even with the Combination Problem looming, Russellian panpsychism can still offer the best available solution to the Mind-Body Problem.

Typically, Conceivability Arguments against physicalism are used to help motivate this conclusion (Chalmers, 1996; 2010; Goff, 2017). However, I shall acknowledge that the Combination Problem can be harnessed into an equivalent Conceivability Argument against Russellian panpsychism (Goff, 2009; Cutter, 2019). In this manner, Conceivability Arguments are of no help to Russellian panpsychism, and, in fact, serve to undermine its motivation further. However, I will propose that *all* philosophy of mind hypotheses are vulnerable to Conceivability Arguments, and then show how their consistent application renders contradictory results. This will lead me to the conclusion that Conceivability Arguments are fundamentally flawed, such that they should be relinquished from both the Russellian panpsychist's repertoire, and indeed, the Mind-Body debate, entirely.

I shall then structure a new argument for Russellian panpsychism and against physicalism, concerning the evolution of consciousness. My argument will show how physicalism struggles to account for consciousness within the evolutionary picture, without adding to the epistemic issues it already suffers. Not only will Russellian panpsychism be shown to be immune to such concerns, but my argument will also serve as a safeguard to Russellian panpsychism, such that even with the Combination Problem looming, Russellian panpsychism can still provide a more attractive theory of consciousness than physicalism. Furthermore, I shall propose that potential future responses to the Combination Problem can be built from the logic of my argument, such that my argument can serve as a foundation for later responses. With these sub-conclusions established, I shall conclude that Russellian panpsychism truly offers the best available solution to the Mind-Body Problem.

In Chapter 1, I will introduce the Mind-Body Problem, alongside the chronology of traditional responses to it. At its core, the Mind-Body Problem reflects the tension that exists between two seemingly ontologically distinct realms: the subjective mental realm, and the objective physical realm (Armstrong, 1999: 1; Crane, 1999; Braddon-Mitchell & Jackson, 2006; Campbell, 1970; Kim, 2010). Prospective solutions to this problem attempt to account for the relationship between the two, in such a manner that each is explained, both in terms of themselves, and in terms of the relations that they share with each other.

There are two main traditional schools of response to this problem: dualism, and physicalism. First, I shall expound substance dualism and property dualism, before explaining how both suffer damning criticism regarding issues of causal closure, and overdetermination (Papineau,

2001; Lowe, 2000; Kim, 1998; 2005; Robinson, 2004). I shall then introduce the currently dominant monistic solution to the Mind-Body Problem – physicalism – and explain how it avoids the issues of dualism, whilst providing a parsimonious potential solution to the Mind-Body Problem. With regards to its own issue, the Hard Problem (Chalmers, 1996: 2010), I shall propose that type-b and type-c variants of physicalism are the most theoretically promising avenues of response. I shall conclude that any theory that wishes to establish itself as possessing the best theoretical solution to the Mind-Body Problem must establish its superiority over type-b and type-c forms of physicalism. This may come in the form of novel explicit arguments against physicalism that weaken its claim further, explicit arguments in favour of the rivalling theory that strengthen its claim over physicalism, or, some combination of the two.

Chapter 2 will then introduce Russellian monism as a potentially viable midground between our traditional theses, such that it appears to capture the virtues of both physicalism and dualism, whilst avoiding their vices. I will explain the contemporary variants of Russellian monism, before settling on a micropsychist form of constitutive Russellian panpsychism as the strongest. I will then present some preliminary considerations about the variant and introduce its main issue: The Combination Problem. I shall conclude that although Russellian panpsychism has many theoretical virtues, as long as its Combination Problem persists, we might not have sufficient motivation to adopt it over type-b and type-c forms of physicalism.

While this thesis will not provide an explicit resolution to the Combination Problem, I explore what reasons we might have to adopt Russellian panpsychism even with this problem looming. In Chapter 3, I will consider one major argument that is typically understood to favour

Russellian panpsychism and oppose physicalism in this vein – David Chalmers’ 2-Dimensional Conceivability Argument (Chalmers, 1996; 2010). Using conceivability, and its alleged links to possibility, the Conceivability Argument alleges to establish that either physicalism is false, or Russellian monism is true. Granting this argument, Russellian panpsychism would offer the best available account of consciousness, even with its Combination Problem unresolved.

In Chapter 4, I will argue that although the Conceivability Argument remains popular, it is of no help to the Russellian panpsychist. In particular, I shall acknowledge that the Combination Problem can be harnessed to produce an analogous argument and conclusion against Russellian panpsychism, such that those who uphold Conceivability Arguments in their support of Russellian panpsychism will find that motivation undermined (Cutter, 2019; Goff, 2009). However, I ultimately find that Conceivability Arguments can be levelled to *all* philosophy of mind hypotheses, such as Russellian panprotopsychism, substance dualism, and property dualism (Goff, 2017; Chalmers, 2016b; Aleksiev, 2023; Brown, 2010; Cutter, 2019). I shall argue that – assuming each argument has equal force – there are three clear possibilities that one may draw. Either: (i) each of my proposed arguments’ premises are true in each application, and that our entire spectrum of hypotheses are incorrect; or, (ii) at least one of their premises is false in each application, and so they cannot provide us with a true conclusion about any philosophy of mind hypothesis; or, (iii) an adequate justification for the truth of at least one of their premises is currently unattainable, and so they are currently of no use within the mind-body debate.

I will then construct a series of valid Conceivability Arguments which negate a metaphysical truth about the world from which they are posited. Building from Katalin Balog criticism of

Conceivability Arguments (1999), alongside Chalmers' subsequent rebuttal (2010), I shall construct three 'Counterfactual Conceivability Arguments' which establish a flawed premise within Conceivability Arguments. This analysis will show that the vulnerability of our philosophy of mind hypotheses to Conceivability Arguments is not necessarily a fault of those hypotheses themselves, but rather a fault within the argument itself. In this manner, I shall discount consequence (i). I shall then locate this fault and propose that consequence (iii) is the most appropriate consequence to draw.

I shall conclude that those who wish to propose Conceivability Arguments as an argument against physicalism will find their own theory vulnerable. However, since I show this to be a fault of Conceivability Arguments, as opposed to a fault within any philosophy of mind hypothesis, I recommend that they are relinquished from the Russellian panpsychist's repertoire. Doing so removes one of the most popularly appealed to arguments for the theory, but such a move is pertinent if the Russellian panpsychist wishes to make a compelling case.

In Chapter 5, I will propose a new argument against physicalism and for Russellian panpsychism. Building from the work of William James (1890) and William Clifford (1874), I will propose a new 'Argument from Evolution,' which shows how physicalism struggles to account for consciousness within the evolutionary picture. My argument begins with two core claims: (1) that the physicalist *should* endorse the precision of consciousness, and (2) that the physicalist who endorses the precision of consciousness will endorse that consciousness is a result of saltation, as opposed to being something fundamental. Upholding these claims, and by working down our evolutionary trajectory with regards to our evolutionary relatives, I will show how the physicalist cannot find any non-arbitrary point where consciousness could have originated aside from the origin of the first living things. I will then show how this conclusion

serves only to worsen the epistemic issues already suffered by physicalism, such that it can no longer be taken to be preferable to Russellian panpsychism.

This final chapter will also serve two further functions. First, it will serve as a safeguard to Russellian panpsychism, such that even with the Combination Problem looming, Russellian panpsychism can still provide a more attractive theory of consciousness than physicalism, provided one avoids attempting to motivate it with Conceivability Arguments. Second, I shall propose that potential responses to the Combination Problem can be built from the logic of my argument, and so this chapter can serve as a foundation for later responses. I shall briefly investigate three potential avenues for the Russellian panpsychist in this regard. After this, I will introduce some historical context with regards to predominant developers of evolution that helps to furnish the conclusion of this chapter. With this in mind, I shall provide Russellian panpsychism with a theoretical ‘safety net’ over physicalism with regards to establishing its internal coherence, and also the grounding to resolve its own internal issues.

1

The Mind-Body Problem

The Traditional Responses ~ Physicalism & Dualism

The focal interest of this thesis is whether Russellian monism provides a better resolution to the Mind-Body Problem than the current dominant approach: physicalism. This chapter will introduce the Mind-Body Problem, the chronology of traditional responses to it, and some relevant terminology. In the first sub-section, I will introduce the Mind-Body Problem alongside what would be required to provide a viable solution. In the second subsection, I will expound the traditional approach of substance dualism, alongside its respective issues: the Interaction Problem, and the Causal Closure Argument. I will then explain the contemporary dualist position of property dualism, before explaining how it too falls to damning criticism. Crucially, while I will refer back to dualism throughout this thesis, it shall not be the direct focus of my analysis. Instead, it is introduced here only to help furnish my analysis with regards to the specific content of physicalism and Russellian monism, alongside how they advance their alleged theoretical solutions to the Mind-Body Problem.

In the third subsection, I will introduce physicalism – the currently dominant monistic solution to Mind-Body Problem. I will explain how it avoids the issues of dualism, whilst providing a parsimonious potential solution to the Mind-Body Problem. I will then focus onto type-b and type-c variants of physicalism as the most theoretically promising physicalist variants and explain how they will form the basis for my comparison of physicalism against Russellian

monism in the remainder of this thesis. I shall conclude that any theory that wishes to establish itself as possessing the best theoretical solution to the Mind-Body Problem must establish its superiority over physicalism. This may come in the form of novel explicit arguments against physicalism that weaken its claim further, explicit arguments in favour of the rivalling theory that strengthen its claim over physicalism, or, some combination of the two. With this in mind, Chapter 2 will introduce Russellian monism, and its theoretical promise as a viable midground between our two traditional theses. The remainder of this thesis will then critically analyse its potential theoretical superiority over physicalism.

1.1 ~ The Mind-Body Problem

Consciousness poses one of the most baffling problems within our scientific understanding of the world (Chalmers, 2010: 1). On the one hand, there is nothing that we are more closely acquainted with than our own conscious experience. On the other hand, how and why it is that we are conscious continues to resist scientific explanation. As physical beings, we have objective properties – such as size, weight, shape, and motion – that are typical of, and have been successfully documented by, scientific inquiry. However, as conscious beings, we also appear to have distinct subjective mental properties that seem atypical of physical entities. Crucially, the latter class of property do not appear to be readily explainable in terms of the former, nor vice versa: they appear to be heterogenous in kind, and they don't seem to fit in the same world (Goff, 2017: 1; Chambliss, 2018: 1). At its core, the Mind-Body Problem reflects the tension that exists between these two seemingly ontologically distinct realms: the subjective mental realm, and the objective physical realm (Armstrong, 1999: 1; Crane, 1999; Braddon-Mitchell & Jackson, 2006; Campbell, 1970; Kim, 2010). Prospective solutions to this problem attempt to account for the relationship between the two, in such a manner that each is explained, both in terms of themselves, and in terms of the relations that they share with each other.

When we refer to the subjective, mental life that we enjoy daily, we see that it is characterised by its abundance of *experience* (Chambliss, 2018: 1). This experience comes in a wide range of modalities, such as visual experience, emotional experience, auditory experience, bodily experience, and even the experience of a conscious thought itself. What unites all of these experiential states, is their inherently subjective nature (Chalmers, 2010: 5). As dramatized by Thomas Nagel in his famous paper ‘What is it like to be a bat?’ (1974), there is ‘something that it is like to be’ an experiencing, conscious thing. For example, there is something it is like to:

- Perceive the redness of a rose.
- Taste the bitterness of coffee.
- Feel the pain of a headache.
- Hear the sound of a violin.

For the purposes of this thesis, we can say that a thing is conscious just in case there’s something that it is like to be it (Goff, 2017: 2). Now, it is important to note that consciousness so defined is not exhaustive of our complete mental life. For example, there exist other aspects of consciousness relevant to higher degrees of cognitive sophistication, including ‘thought,’ ‘awareness’ and those other psychological attitudes towards specific propositions that one enjoys as a higher-order conscious thing. For example, one might think about their perception of the redness of a rose in terms of previously experienced roses, or in terms of how that rose makes one feel at that moment. In order to distinguish the relevant notion from those more cognitively complex, I shall henceforth refer to the ‘something it is like to be’ a conscious thing as ‘phenomenal experience.’

My reasoning for distinguishing these aspects is threefold. First, phenomenal experience is a hard datum that any adequate theory of reality must accommodate: I know that there is something it is like to be me; you (the reader) know that there is something that it is like to be you; and we probably both think there is something that it is like to be a hamster (Goff, 2017: 3). Second, phenomenal experience is a clear part of our common-sense picture of the world and is not excessively restrictive in its parameters of application. For example, while we probably both think there is something it is like to be a hamster, we might not wish to attribute those more cognitively sophisticated aspects of our complete mental life to it (Goff, 2017: 2). Third, and most crucially, the ‘something it is like to be’ a conscious thing might be the most crucial aspect to be accounted for in our explanation of consciousness (Lee, forthcoming: 12). By virtue of it being a less restrictive term, and since those more cognitively complex aspects of consciousness are still represented through our experience *of* them, phenomenal experience appears to be a foundational aspect of all consciousness. For example, when we think about our perception of the redness of a rose, our thought is directed upon and represented through our experience of the rose.¹ In this manner, experience appears to precede these other aspects of consciousness – such as thought – in such a way that once experience is accounted for within our understanding of the world, those higher order forms of cognition might be explained away. For these reasons, ‘phenomenal experience’ seems to be the most logical starting point to any explanation of consciousness and the relation it stands in with regards to the physical world.

With this in mind, we can define the relevant aspect of consciousness as:

¹ While this may be uncontentious to the reader, it is worth noting that some authors do contest the reality of this sort of ‘cognitive phenomenology,’ and instead, think that it can be reduced entirely to sensory phenomenology. See Prinz (2011: 174-196) and Tye and Wright (2011: 326-344) for arguments in this vein.

Phenomenal Experience_{ar}: The first-person ‘what it is like to be’ aspect of conscious experience.

With this clarification in place, we can say that the core of the Mind-Body Problem is the problem of characterising and explaining phenomenal experience and the relation it stands in with regards to the physical world (Armstrong, 1999: 1; Crane, 1999; Braddon-Mitchell & Jackson, 2006; Campbell, 1970; Kim, 2010). Or, more formally, we can propose that the Mind-Body Problem highlights the following requirement:

The Consciousness Constraint_{ar}: Any adequate theory of reality must entail that phenomenal experience is explained.

(Goff, 2017: 3)

Clearly, any adequate theory of reality should include an explanation of and for the mental life of conscious organisms. The issue as it stands, however, is that solely empirical approaches do not seem able to do this. Over the last half millennia, our scientific theory of reality has been extremely fruitful by virtue of its application of methodological naturalism (Ladyman et al, 2007; Churchland, 2013: 57-8). Within this approach, one looks to the third-person scientific method, viz. rigorous empirical investigation of the publicly observable, to tell us what reality is like. The application of this approach is then subsequently reinforced through its contemporary and historical success in the natural sciences (Goff, 2017: 5). The issue highlighted by the Mind-Body Problem, however, is that is that the defining feature of phenomenal experience is its inherent subjectivity. The ‘something it is like to be’ a conscious thing, is inherently personal; it has a ‘for-you-ness’ that is accessible only to the subject who is experiencing it. Since empirical data are, by definition, publicly assessable and

intersubjectively verifiable, phenomenal experience appears to be an essentially *non-empirical* datum. With this in mind, our historically successful approach of methodological naturalism appears unable to help us in characterising and explaining phenomenal experience.

Now, it is generally clear to everyone that everything that happens in the mind is closely linked to whatever is happening in our brains and bodies. However, since the defining feature of phenomenal experience is its inherent subjectivity, phenomenal experience seems incomprehensible in terms of our brain's physical constitution or operation. Crucially, this problem is not due to a lack of empirical information about our brains, bodies, or the physical constituents of the universe: the problem is theoretical in kind, not empirical (Nagel, 2007). For instance, the solid, extended properties of brain matter appear entirely distinct from the seemingly un-extended, subjective properties of mind. Furthermore, our knowledge of either property comes from necessarily distinct forms of inquiry; the neural processes of the brain are known solely through third-person empirical investigation, and the subjective first-person perspective of the mind is known solely through introspection (Goff, 2017: 1).

Because of this enduring tension between the subjective mental realm and the objective physical realm, there is no experiment that could tell us what is required from reality so as to satisfy the Consciousness Constraint (ibid: 9). For this reason, working out what is required to solve the Mind-Body Problem is a *philosophical*, not a *scientific* task. This is not to say that we must necessarily add to the current conception of reality that we uphold as standard in order to satisfy this requirement. For example, it could turn out that all of the current physical things postulated to account for the third-person empirical data studied by the methodological naturalist happen to be sufficient to also account for the reality of consciousness (ibid). The point is that this cannot be assumed, and that science alone cannot tell us whether or not this is

the case. In this regard, and for the remainder of this chapter, I shall introduce two separate schools of response to the Mind-Body Problem, each of which holds different ontological commitments about the nature of reality, and consciousnesses place within it.

1.2 ~ Dualism

Broadly speaking, dualism is the thesis that the mind and body are two very different forms of thing (Armstrong, 1999: 9). In this subsection, I will expound two of its most traditionally popular variants: substance dualism, and property dualism. I will explain the key features of both theses with regards to the Mind-Body Problem, alongside their respective criticisms: the Interaction Problem and Causal Closure Argument against substance dualism, and the claim that property dualism either results in overdetermination or causal redundancy. I shall conclude that by virtue of these problems, substance dualism does not provide a viable solution to the Mind-Body Problem, and property dualism either forces epiphenomenalism or reduces into physicalism. In this manner, I shall find that neither variant provides a viable solution to the Mind-Body Problem, and so turn to the more commonly accepted doctrine of ‘physicalism’ in the next subsection.

Of the two variants that I will expound here, **substance dualism** has been the most traditionally adopted. Heavily popularised by René Descartes in the 17th century, substance dualism intends to preserve the distinctiveness of both mental and physical phenomena, while maintaining their causal interaction within us as unified human beings (Cottingham, 1992: 115). Substance dualism proposes that the mind is a distinct mental substance in-and-of-itself, but which can nevertheless interact with the body despite their ontological differences. We can understand ‘substance’ here as a “thing which exists in such a way as to depend on no other thing for its existence” (Descartes, 1985a: 25). In order to establish this ontological distinction, Descartes

proposes three arguments. Each will appeal to the Indiscernibility of Identicals, or ‘Leibniz’s Law:’ the principle that if two entities are identical, then they must share all of the same properties (Leibniz, 1989: 216). More formally, we can put this principle as follows:

Leibniz’s Law *df.* If x and y differ with respect to any property P , then $x \neq y$.

Or,

$$\forall x \forall y [(x=y) \Rightarrow (\forall P (P(x) \Leftrightarrow P(y)))] \quad \text{and} \quad \exists P [P(x) \wedge \neg P(y)] \Rightarrow (x \neq y)$$

Descartes’ first argument – The Argument from Doubt – employs his infamous method of Cartesian doubt about the reliability of sensory perceptions and the existence of the material world (ibid: 1985: I, 18). Descartes proposes that it could be the case that one is being deceived by, for example, an evil demon, who manipulates our perceptions such that we incorrectly view the external world (ibid: I, 22). In contrast, he proposes through his famous ‘cogito, ergo sum,’ that our mind is indubitable, as even if all our other perceptions and beliefs are false, the very act of our doubting confirms the existence of the doubter (ibid: II, 25). In this manner, he concludes that one can doubt the existence of the body, but not the existence of the mind. Following Leibniz’ Law, since body and mind have different properties, they cannot be one and the same.

Descartes’ second argument – The Argument from Divisibility – builds from the apparent unity of our consciousness. Descartes proposes that when he considers the mind, he sees that it lacks any special extension, and that he cannot distinguish any parts within it: it is a simple and complete indivisible whole (ibid: VI, 85). By contrast, he proposes that since the body is extended in space, it can be divided into parts: by its very nature, body is always divisible (ibid). For analogy, while it makes sense to speak of ‘half a brain,’ or ‘half a liver’, and so on, it intuitively makes no sense to speak of ‘half of the feeling of joy,’ ‘half of a headache,’ or

‘two-thirds of a belief.’ Now, since divisibility may be predicated of bodies, but not minds, then per Leibniz’s law, minds cannot be identical to bodies.

Descartes final argument – The Argument from Clear and Distinct Ideas – reinforces this logic by claiming that a *real distinction* exists between mind and body. Descartes proposes that by virtue of God’s existence as a non-deceptive entity, whatever one can clearly and distinctly perceive to be true must be true (ibid: III, 35). He argues that we have a clear and distinct idea of our minds as thinking, non-extended substances (ibid: II, 27), alongside a clear and distinct idea of our bodies as non-thinking, extended substances (ibid: VI, 78) Since clear and distinct ideas are necessarily true for Descartes, then mind and body *must* have different properties. Following Leibniz’s Law again, if we can clearly and distinctly conceive of two substances as having different properties, then they must then be ontologically distinct substances (ibid).

To illustrate this more succinctly, we can say that x is body, y is mind, and property P refers to either dubitability, divisibility, or extension. It follows from Descartes’ three arguments that $\exists P [P(x) \wedge \neg P(y)] \Rightarrow (x/y)$; x has the property of dubitability, divisibility, or extension, where y does not, and so x cannot be the same as y . Similarly, if property P refers to either indubitability, indivisibility, or non-extension it follows that $\exists P [P(y) \wedge \neg P(x)] \Rightarrow (y/x)$; y has the property of indubitability, indivisibility, or non-extension, where x does not, and so y cannot be the same as x . Following Leibniz’s Law then, Descartes allegedly establishes that mind and body must be distinct.

Despite mind and body being distinct substances, Descartes believes that they can still causally interact with one another and form a union into a single integrated human being. For example, he proposes that he is not only lodged in his body as a pilot in a ship, but that he is

also very closely joined and, as it were, intermingled with it (ibid: VI, 81). To account for this interaction, Descartes describes the pineal gland as the ‘locus’ where the mind and body converge: “The small gland which is the principal seat of the soul... is suspended within the cavities containing these spirits, so that it can be moved by them and can in turn move them” (Descartes, 1985c: 354). This site of their union is alleged to enable the mind to experience bodily sensations, such as heat, hunger, or thirst, which cannot be understood solely in terms of their mental or physical components, and also for the mind to causally influence the body of its will.

Descartes’ substance dualism attempts to preserve the distinctiveness of the mind and the body, whilst also allowing for their interaction. It alleges to resolve the Mind-Body Problem by proposing the mind – and so phenomenal experience – exists as a distinct substance. Similarly, by preserving the distinctiveness of phenomenal experience in this way, it also appears to satisfy the Consciousness Constraint. However, substance dualism is not commonly accepted as a viable solution to the Mind-Body Problem due to a number of heavily debated problems. At the forefront of these, are the Interaction Problem, and the Causal Closure Argument.

The Interaction Problem concerns the difficulty of explaining how two radically different substances could influence or be influenced by each other. For Descartes, mind is non-spatial and indivisible, while body is spatially extended and divisible. The mind is then proposed to cause bodily movements through volitions, and the body is proposed to influence the mind through sensory inputs. However, it is unclear how two substances with entirely different natures could engage in the required causal relations for this sort of interaction to occur. One of the earliest proponents of this criticism was Princess Elisabeth of Bohemia. In her correspondence with Descartes, she writes: “I beseech you to tell me how the soul of a human

being... can determine the movement of the animal spirits in the body so as to bring about voluntary actions” (Elisabeth to Descartes, May 1943: 611).

Elisabeth’s critique highlights a clear conceptual gap in Descartes’ explanation: causation appears to require a shared property (e.g. spatial extension) that mind and body simply do not have. For example, physical force is necessary to cause bodily movement, but if the mind is non-extended, it cannot exert any physical force. In this manner, Descartes appeal to the pineal gland as the site of mind-body interaction does nothing to actually explain that interaction. While Descartes acknowledges this issue in later correspondence with Elisabeth, he provides no explicit resolution to the problem. Instead, he proposes that although the manner of interaction is incomprehensible, it occurs as a matter of fact, nonetheless (Letter to Elisabeth, May 28: 665).

The Causal Closure Argument provides a more modern and damning extension of the traditional Interaction Problem. The Principle of Causal Closure of the physical domain states that:

1. All physical events have sufficient physical causes.
2. No non-physical cause is required to explain physical phenomena.

(Papineau, 2001: 17).

For example, when we strike a pool ball with a cue, and we see the ball roll, the direction of movement, rotation, and velocity of the ball in this situation is entirely accounted for by its impact against the pool cue. In this manner, the event of the ball rolling has an entirely sufficient physical cause, such that we could predict exactly where the ball will end up, and how long it will move for, with a knowledge of the physical variables involved. In this manner, we do not

have to appeal to any other cause to account for the phenomena of its rolling. With this in mind, the Principle of Causal Closure is then inputted into arguments of the form of:

- P.1 At every time at which a physical state has a cause, it has a fully sufficient physical cause (Principle of Causal Closure).
- P.2 Some physical states have mental states amongst their causes.
- P.3 When a physical state has a mental state amongst its causes, it is rarely if ever causally overdetermined by that mental state and some other physical state.
- C. At least some mental states are identical with certain physical states.

(Lowe, 2000: 27).

This argument proposes that if physical events are fully caused by prior physical events, the addition of a mental cause would lead to casual overdetermination – where the effect has two distinct and sufficient causes (Kim, 1998: 41-44). For instance, if the movement of an arm is entirely explained by physical events in the brain and the body, invoking a mental decision as an additional cause becomes redundant and metaphysically implausible.² In this manner, one cannot reconcile immaterial mental causation with the causal completeness of the physical domain, and so the motivation for substance dualism is undermined. This, when combined with the interaction problem provides a damning criticism of substance dualism: since a mental substance cannot intelligibly interact with the physical domain, and since events in that physical domain are already fully accounted for by physical causes, then appears little reason to propose a distinct mental substance in the first place.

² While potential responses to this problem do exist, cf. Gibb (2015). I shall omit them here for the sake of brevity.

Because of these sorts of issues, substance dualism has broadly fallen out of favour. **Property dualism** has since emerged as an attempt to resolve the above issues while retaining the essence of dualism (Chalmers, 1996; Jackson, 1982). In contrast to substance dualism, property dualism maintains an ontological *monism*, whereby there is only one fundamental type of substance in the universe, and it is physical. In this manner, it intends to avoid the respective issues of substance dualism regarding the interaction of two distinct ontological realms. However, it also maintains property *pluralism*, such that it posits two fundamentally distinct types of properties: physical properties and mental properties (of which phenomenal experience is one). Crucially, these mental properties are irreducible to physical properties, such that they cannot be fully explained in physical terms, but nevertheless supervene on physical properties such that any change in mental property incurs a corresponding change in physical property (O'Rourke, 2018; Vintiadis, 2019; Robinson, 2004: 42). In this light, property dualism attempts to reconcile a dualistic account of mind and body with a monistic account of substance, and which allegedly accounts for the interaction between these two classes of property (Kim, 1993: 53).

The main motivation for property dualism is to provide a solution to the Mind-Body Problem which satisfactorily accounts for the subjective nature of mind and avoids the issues and substance dualism whilst still acknowledging the potential limits of explanation provided by the natural sciences. For example, property dualists propose that while neuroscience can explain the neural correlates of consciousness, it cannot explain why or how these neural processes are experienced subjectively (Chalmers, 1996: 4). In this manner, while physical explanations of brain activity can tell us how neurons interact, they cannot tell us anything about phenomenal experience. For this reason – and holding onto that same intuition that motivated substance dualism – property dualists propose that consciousness must be a non-

physical property of the brain that cannot be entirely reduced to its physical processes (Jackson, 1982: 127-129).

Although property dualism provides a new way of acknowledging phenomenal experience without running into the explicit issues involved in invoking separate substances, it faces several powerful challenges, nonetheless. The most potent of these challenges is a similar interaction problem to the substance dualist. In particular, the property dualist must answer the question of how it is that mental properties can interact with physical properties in a causal capacity (Robinson, 2004). Now, per the causal closure principle, every physical effect has a sufficient physical cause (Kim, 1998: 40). However, if a physical effect has a sufficient cause, then there is no need to posit an additional distinct cause for the same effect, as no single event typically has more than one sufficient cause (ibid, 2005: 42). Because of this, mental properties appear to be either causally inert, or epiphenomenal. For example, if some physical effect (P2) is sufficiently caused by a physical event (P1), and a mental event (M1) is posited as an additional cause of P2, then P2 is causally overdetermined, *unless* M1 is reducible to P1. To avoid such overdetermination, one must exclude M1 as a cause, rendering mental causation epiphenomenal. In this manner, we can say that the positing of distinct mental properties will result in either overdetermination, or causal redundancy (Kim, 2005: 42; 1998).

This problem appears to pose a difficult dilemma for property dualism. The first option is to accept that mental properties are causally inert, which reduces property dualism to an epiphenomenal claim that no longer supports the dualist intuition or purpose. The second option is to respect the causal closure of the physical domain and accept that mental properties are reducible to physical properties, such that they do not supervene, and therefore be led to a

reductive physicalist approach (Kim, 2005: 76).³ Either way, property dualism appears to become impotent, and so is not commonly accepted as a viable solution to the Mind-Body Problem.

1.3 ~ Physicalism

By the mid-20th century, dualist views were broadly rejected in favour of more empirically supported monistic views, such as physicalism (or, materialism). For instance, J.J.C. Smart's seminal paper, *Sensations and Brain Processes* (1959: 141-156) argued that mental states are nothing over and above brain states, and this claim initiated a widespread rejection of non-physicalist theories that has broadly continued until today. For the purposes of this thesis, we can understand physicalism to be a monistic thesis which proposes that everything that exists is fundamentally physical in nature (Goff, 2017: 23). In its purest form, this simply means that there is only one fundamental substance – the physical – and that its reality can be entirely captured and accounted for by the language of physics; physical facts constitute the fundamental nature of the world.⁴ These include facts about physical entities, properties, relations, and laws (including those known, and those currently not known but which nonetheless exist throughout space and time) that can be captured in a mathematico-nomic vocabulary (Goff, 2017: 31; 79-80). For physicalism, there is no duality in one's ontology of substance or property; consciousness, like everything else in the universe, is explained by the physical.

We can formulate the following preliminary definition for the doctrine of physicalism:

³ This approach will be the focus of the next subsection.

⁴ It is worth noting that not all physicalists define physicalism in this way. For example, according to Alyssa Ney (2008: 1033–48), physicalism does not name a metaphysical view, but rather an attitude to metaphysical enquiry.

Physicalism_{af}: Fundamental reality is wholly constituted of facts that (i) concern spatio-temporal entities at relatively low levels, that can be captured in the vocabulary of physics, and which (ii) do not involve phenomenal experience (Goff, 2017: 40).⁵

The primary motivations for physicalism are threefold. The first is due to the vast and historical explanatory power of the physical sciences in describing and predicting phenomena. As Daniel Stoljar notes, "the development of physical science has been the outstanding intellectual achievement of the last five hundred years" (Stoljar, 2010: 1). Historically, as science has advanced, phenomena previously considered non-physical – for example, life processes – have come to be explained away in physical terms. This success suggests that a physicalist framework is well-suited to come to account for phenomenal experience, even if it cannot yet. Furthermore, consciousness is clearly deeply intertwined with physical processes, and advances in neuroimaging have correlated many specific mental states with neural activity. For example, the localisation of visual processing in the occipital lobe and the association of emotional processing with the amygdala highlight how alterations in brain structure and activity produce predictable changes in mental states (Churchland, 1986: 50). In this manner, it might be thought that the progress of physical science provides inductive evidence that phenomenal experience will also come to be explained away in physical terms (ibid: 21).

The second motivation is due to the aforementioned principle of causal closure, and dualism's related issues regarding it. If we are to accept the principle of causal closure – which we appear to have good reason to – then we need not ever appeal to anything non-physical (Papineau,

⁵ While there do exist other classifications within physicalism – such as emergentist forms (cf. Broad, 1925; Alexander, 1920; Deacon, 2012) or functionalist forms (cf. Putnam, 1967) – they face similar criticisms to those proposed of property dualism, and their particular claims are outside of the purview of this thesis.

2002: 17). If no non-physical causes are needed to account for physical phenomena, and if physical events are to be influenced by mental events, then mental events must themselves be part of the physical causal nexus (ibid: 32). In this manner, through making phenomenal experience physical, the physicalist avoids positing unexplained influences outside of the physical domain, and instead, integrates consciousness into it. In doing so, physicalism unifies consciousness with science, and avoids the aforementioned issues suffered by dualism.

Finally, one can motivate physicalism simply by an appeal to simplicity. On the one hand, physicalism avoids the ontological complexity of positing a separate realm for mental substances. By reducing all phenomena – including phenomenal experience – to the physical, it satisfies Occam's Razor, which states that entities should not be multiplied beyond necessity. On the other hand, it also provides a greater conceptual unity: its singular and unified framework exists to explain *all* phenomena, whereas dualism fractures into physical and non-physical domains. In this manner, physicalism appears to offer the more parsimonious account of all phenomena, and eliminates any unnecessary metaphysical complications (Churchland, 1986: 46).

With this broad definition of physicalism set up, alongside the reasons for its adoption, let us turn to the question of what it would mean for fundamental reality to be entirely physical. This calls for an account of 'fundamentality,' or what it is for certain (in this case physical) facts to entirely constitute fundamental reality. Now, fundamentality is typically understood in terms of grounding. Broadly speaking, grounding refers to the non-causal relation that connects more fundamental entities or facts to less fundamental ones, thereby explaining how or why certain phenomena or truths exist. It assumes an ontological hierarchy, where the less fundamental is reducible to or dependent upon on the more fundamental. Generally speaking, we can say that

if some fact X grounds fact Y, then Y is nothing over and above X (Goff, 2017: 42; Fine, 2001: 1-23).

It will be useful here to appeal to an example to help illustrate this notion. Suppose Jack, Jill, and James are drinking, dancing, and generally being merry one evening at Jill's. It follows from this fact that there is a party at Jill's. Moreover, it follows that there is a party *because* Jack, Jill, and James are drinking, dancing, and so on. This is not a causal relationship; the actions of Jack, Jill, and James does not bring the party into being. Rather, the party is simply nothing over and above their actions. Along similar lines, we can consider a scarlet-coloured hydrangea flower. It follows from this fact that the hydrangea is red *because* it is scarlet. Again, the scarlet colour of the hydrangea does not *cause* the flower to be red. Rather, the hydrangea's being red is simply nothing over and above its being scarlet (Goff, 2017: 42). In both cases, the grounded entity is nothing over and above its ground, rather than the ground causing the grounded entity.

For the purposes of this thesis, we shall adopt Philip Goff's way of understanding grounding, named 'grounding by analysis' (Goff, 2017: 44).⁶ Within this account, we can say that some fact Y is grounded by analysis in fact X iff:

- Y is grounded in X, and
- X logically entails what is essentially required for the entities contained in Y (including property and kind instances) to be part of reality.

(Goff, 2017: 45).

⁶ I adopt this notion due to it providing a clear and concise definition of a popular idea within the literature (such as that presented by Dasgupta (2014), or Melnyk (2003)).

In reference back to the aforementioned party example, we can say that the fact (Y) that there is a party is grounded by analysis in the fact (X) that Jack, Jill, and James are drinking, dancing, and generally being merry because:

- There being a party at Jill's is nothing over and above the fact that Jack, Jill, and James are drinking, dancing, and generally being merry, and
- The fact that Jack, Jill, and James are drinking, dancing, and generally being merry logically entails what is essentially required for there to be a party.

(ibid; Dasgupta, 2014; Melnyk, 2003).

Grounding by analysis exists here, due to the nature of what a party is. Intuitively, a party is the kind of thing in which a collection of people dance, drink, and are generally merry. In this manner, the nature of what a party is 'points' to its possible grounds, such that 'a party' is open to the possibility of being grounded in specific facts concerning people drinking, dancing and being merry (Goff, 2017: 49-50; Fine, 2012). Crucially, however, grounded facts need not necessitate their grounds: it isn't the case that a party must always be grounded in Jack, Jill, and James drinking, dancing, and being merry. However, it is generally accepted that there is a necessitation relation going in the other direction, such that any grounded fact is necessitated by the obtaining of its ground.⁷ In this manner, if Jack, Jill, and James *are* drinking, dancing, and being merry, then there is a party *necessarily*:

Necessitation_{ar}: $X_1 \dots X_n$ necessitates Y when the conditional, 'if $X_1 \dots X_n$ then Y' is metaphysically necessary.

⁷ Although this assumption is generally accepted, Barbara Montero (2013), for example, has questioned it.

In this manner, if we can say that if some facts $X_1 \dots X_n$ ground fact Y , then it is metaphysically necessary that if $X_1 \dots X_n$, then Y . By this logic, any grounded fact or entity is necessitated by the obtaining of its ground, such that if there is a world in which X , then Y , necessarily (Cutter, 2018: 45).

With this set up, we can now input our grounding account into the claims of physicalism. Since physicalism claims that there is only one fundamental substance – the physical – then that substance must ground and explain all other (grounded) entities. Accordingly, we can input this account in a more robust definition of physicalism that shall be adopted in this thesis:

Physicalism_{df}: All facts that concern entities at greater than fundamental levels are grounded by analysis in facts about fundamental physical entities whose nature (i) can be captured in the vocabulary of physics, and (ii) does not involve phenomenal experience.

With this in mind, we can say that physicalism's response to the Mind-Body Problem is that consciousness is grounded by analysis in facts about fundamental physical entities. Now, if physical facts ground and so necessitate the reality of consciousness in this way, then any minimal physical duplicate world to ours – viz., one which contains the exact same fundamental arrangement of physical parts – will contain consciousness, necessarily (Goff, 2017: 79). Consciousness would be a higher-order physical process that is dependent upon lower-level physical parts and processes for its occurrence (Chalmers, 2010: 105). In this manner, the physicalist proposes that phenomenal experience is entailed by its ground viz., in terms of fundamental physical parts and processes.

While physicalism finds support through the vast and historical explanatory power of the physical sciences, its avoidance of dualism's concerns, and, indeed, its general parsimony, it remains susceptible to a number of problems with regards to explaining consciousness. These can be delineated into '**easy problems**' and '**The Hard Problem**' (Chalmers, 1996). The easy problems are those that seem directly susceptible to the standard methods of cognitive science, whereby a phenomenon is explained in terms of neural mechanisms. For example, the physicalist must account for the following phenomena: the focus of attention, the control of behaviour, the difference between wakefulness and sleep, and so on. Although we do not yet have anything close to an explanation for these sorts of phenomena, there is no real theoretical issue in explaining them scientifically: we need only specify the relevant neural mechanisms that can perform those functions.

The Hard Problem, on the other hand, is the problem of experience itself. As noted previously, the Mind-Body Problem highlights the Consciousness Constraint, whereby any adequate theory of reality must entail that phenomenal experience is satisfied. While physicalism claims phenomenal experience to be grounded in facts concerning fundamental physical entities, they then open themselves to the problem of explaining how and why it is that physical parts and processes give rise to phenomenal experience in the first place (Chalmers, 2010: 105, 5). In this manner, we can say that the **Hard Problem** is the problem of accounting for:

Why and how phenomenal experience arises from physical processes in the brain.

Unlike the 'easy problems' which simply involve explaining cognitive and behavioural functions in terms of neural mechanisms, the Hard Problem focuses on the ontological nature

of phenomenal experience itself.⁸ For example, why is it that the activation of certain neural circuits results in the vivid sensation of ‘redness’, the pain in my lower back, or the fragrant smell of coffee? (ibid: 8). While neuroscience is theoretically capable of providing a detailed or complete account of the brain’s functional process, it arguably lacks the tools – and, perhaps, the intent – needed to bridge the epistemic gap between objective brain activity and inherently subjective phenomenal experience. For instance, no matter how much we discover about some entity’s nervous system, we will not discover what the experiences had by that entity were *like* (Robinson, 1982: 4).

It appears that no amount of new information about the physical processes of a conscious thing will amount to information about *what it is like* to be that thing. In this manner, there is a clear sense in which this problem may persist even after the performance of all the relevant neural mechanisms (and easy problems) are explained. In fact, due to its focus on solely physical forms of explanation, physicalism appears to distort and disregard the nature of our conscious experience in such a manner that everything we learn from its explanation is consistent with an absence of conscious experience entirely (Goff & Coleman, 2020: 301). With this considered, and in light of the Consciousness Constraint, if physicalism cannot account for the Hard Problem, then it cannot provide a viable solution to the Mind-Body Problem.

There are roughly three ways that a physicalist might resist or respond to the Hard Problem. Each variant offers a different theoretical response, that, if vindicated, would remove the problem. The first of these is through ‘**type-a physicalism.**’ Now the type-a physicalist simply denies that there is an epistemic gap between physical and phenomenal truths. In this manner, they maintain that there is no Hard Problem of explaining consciousness that will remain once

⁸ ‘Functions’ should be read here as any causal role in the production of behaviour that a system might perform.

the easy problems have been resolved (Chalmers, 2010: 111).⁹ More recently, this response takes the form of ‘eliminativism,’ which holds that consciousness does not exist, in such a manner that there are no phenomenal truths (Churchland, 1986; Churchland, 1984; Rorty, 1979). It can also take the form of analytic functionalism, or logical behaviourism, which hold that consciousness exists, but where the concept of ‘consciousness’ is defined wholly in functional or behavioural terms, respectively (Putnam, 1967; Fodor, 1975; Ryle, 1949; Feigl, 1962). The characteristic feature of all of these views is that there is nothing about consciousness that needs explaining over and above the various functions that constituted the easy problems of consciousness.

The **type-b physicalist** responds to the Hard Problem by claiming that although there is a genuine epistemic gap between the physical and phenomenal domains, it is not an ontological gap.¹⁰ In this manner, while there is a Hard Problem distinct from the easy problems, it does not correspond to a distinct ontological domain (Chalmers, 2010: 115). Typically, this form of response maintains that phenomenal states can be identified with certain physical or functional states, in the same manner that an identity exists between water and H₂O, or between genes and DNA. However, unlike in these cases – where the identity between them is deducible from the complete physical truth about the world – the type-B physicalist maintains that the identity between consciousness and physical or functional states is *epistemically primitive*, such that the identity is not deducible from the complete physical truths (ibid: 116). Indeed, if it was, then type-a physicalism would be correct instead. Along these lines, the type-b physicalist

⁹ Prominent type-a physicalists include Daniel Dennett (1991), Fred Dretske (1995), Gilbert Harman (1990), David Lewis (1988), Georges Rey (1995), and Gilbert Ryle (1949).

¹⁰ Prominent type-b physicalists include Ned Block and Robert Stalnaker (1999), Christopher S. Hill (1997), Joseph Levine, (1983), Brian Loar (1990), William Lycan (1996), David Papineau (1993), John Perry (2001), and Michael Tye (1995).

maintains that although we will never be able to conceptually deduce the identity between certain physical and phenomenal states, the two share identity, nonetheless.

Finally, **type-c physicalism** responds to the Hard Problem by claiming that although there is currently a deep epistemic gap between the physical and phenomenal, the gap is closable in principle. In this manner, the Hard Problem is in principle solvable in physical terms, such that a complete knowledge of the physical truths will show there to be no ontological gap between the domains, whilst also removing the current epistemic one. The most common argument for this response is that as humans, we have cognitive and epistemic limitations: we are limited by our own minds, or by the current incompleteness of the physical sciences. For example, Thomas Nagel (1974), has suggested that just as how the pre-Socratics could not have understood how matter could be energy, so too can we not yet understand how consciousness could be physical. Similarly, Patricia Churchland (1997), has suggested that we are simply burdened with a psychological limitation that further scientific progress will overcome. Finally, Colin McGinn (1989) has suggested that the epistemic problem may be unsolvable by humans due to limitations in our cognitive abilities, but that it, nevertheless, has a solution in principle.

Physicalism, in each of its many variants, is the dominant theoretical solution to the Mind-Body Problem within contemporary academia. As noted, this is primarily due to its alignment with empirical findings in neuroscience, its comprehensive explanatory framework, and the historical success of the physical sciences. In this manner, physicalism offers a parsimonious and unified account of the world which proposes to eliminate the need for non-physical entities or dualistic accounts of substances. Furthermore, the explanatory unification it offers with regards to the methodological naturalism that underpins the empirical sciences represents a significant virtue that reinforces its epistemic credibility as a doctrine (Churchland, 1986: 52).

While physicalism could yet turn out to be false by virtue of its Hard Problem, any theory that wishes to establish itself as possessing the best theoretical solution to the Mind-Body Problem must establish its superiority over physicalism. This may come in the form of explicit arguments against physicalism (in addition to those already expounded) that weaken its claim further, explicit arguments in favour of the rivalling theory that strengthen its claim over physicalism, or, some combination of the two.

Henceforth, only the type-b and type-c forms of physicalism shall be of interest to this thesis. I exclude type-a solely due to many of its forms of response falling outside the purview of what could constitute a viable solution to the Mind-Body Problem. For example, eliminativism appears to deny the manifest in such a manner that it simply denies the Mind-Body Problem. Similarly, analytic functionalism appears to amount to the claim that there is no Hard Problem of consciousness over and above those easy problems; rather than confronting the problem, it simply redefines it. In this manner, there is a clear sense in which many type-a forms of response do not register phenomenal experience as an explanandum in the relevant sense and so violate what is requested by the Consciousness Constraint (Chalmers, 2010: 112). Now, while type-a responses may yet turn out to be legitimate, I continue under the assumption that phenomenal experience exists, and that there is a Hard Problem. With this in mind, the focus of my analysis regarding physicalism within this thesis will be on its type-b and type-c forms instead.

1.4 ~ Final remarks

In this chapter, we have introduced the Mind-Body Problem, alongside two of the main traditional approaches to solving it: physicalism and dualism. First, I expounded the traditional approach of ‘substance dualism,’ before delving into the main problems that it faces: the

Interaction Problem and the Causal Closure Argument. I then introduced its more contemporary cousin – property dualism – before explaining how it too falls to similar criticism that appears to render its thesis impotent. With that in mind, we turned to the dominant monistic solution to the Mind-Body Problem: physicalism. I explained how it avoids the issues of dualism, whilst providing a parsimonious potential solution to the Mind-Body Problem, but which nevertheless falls victim to the Hard Problem.

We can say that physicalism has the virtue of integrating consciousness into the natural order and is capable of explaining many kinds of mental phenomena. However, it also appears to fail to explain the existence of consciousness itself. We can say then that physicalism has the virtue of integrating consciousness into the natural order, but the vice of the Hard Problem. Dualism, on the other hand, has the virtue of avoiding the Hard Problem: by positing consciousness as something fundamental in-and-of-itself, it does not distort or disregard its nature when attempting to explain it. However, it has the converse vice of separating consciousness from the natural order and so is potentially incapable of integrating consciousness within physical causation and of resolving issues of overdetermination.

Notwithstanding these respective issues, I focussed onto type-b and type-c variants of physicalism as the most potentially theoretically promising physicalist solutions to the Mind-Body Problem: they were shown to employ a comprehensive and historically successful explanatory framework, and also to deliver a theoretical solution to the Hard Problem which could satisfy the Consciousness Constraint in the appropriate way. I explained that these variants will form the basis for my comparison of physicalism against Russellian monism for the remainder of this thesis. With this in mind, if we are to find Russellian monism more theoretically promising than physicalism with regards to the Mind-Body Problem, then one

must level further explicit arguments against physicalism, explicit arguments in favour of Russellian monism that strengthen its claim over physicalism, or, some combination of the two.

2

Russellian Monism

A Potential Midground Solution

‘Russellian monism’ denotes a collection of theories primarily inspired by Bertrand Russell’s discussion of physics within *The Analysis of Matter* (1927a) and other successive texts (1927b; 1948; 1959; 1984). Rather than representing the theory to which Russell himself subscribed, Russellian monism simply refers to a family of views that are united with and inspired by his work. While Russell is the doctrine’s namesake, his contemporary Arthur Eddington was also formative to its creation. Eddington expressed a complementary idea to Russell in *Space, Time, and Gravitation* (1920), and *The Nature of the Physical World* (1927). While many aspects of this work were positively influenced by Russell, some influenced Russell respectively. For this reason, I shall appeal to both authors in my exposition of the doctrine.¹¹ The doctrine itself did not receive much attention throughout the latter half of the twentieth century. However, the work of numerous contemporary philosophers, alongside neuroscientists and physicists has brought about its re-emergence as a serious metaphysical theory – one in which I shall argue may hold the greatest theoretical promise for resolving the Mind-Body Problem.

In the first section, I will set out the traditional formulation of the doctrine – as inspired by the work of Russell and Eddington – and explain the theoretical midground and potential virtues that this doctrine provides over both physicalism and dualism with regards to resolving the

¹¹ Russellian monism also holds similarity with some of the foundational elements of Alfred North Whitehead’s process-philosophy (1929a; 1929b; 1938).

Mind-Body Problem. In the second section, I will introduce the main contemporary dimensions of Russellian monism and clarify the particular variant which I believe to be most promising: a constitutive and micropsychist form of Russellian panpsychism. I will then present some preliminary considerations about this variant and introduce the main issue suffered by the theory: the Combination Problem.

Russellian panpsychism appears to capture the virtues of both physicalism and dualism while avoiding their vices. In this manner, if it can resolve its Combination Problem, then it would become the best available theory of consciousness by default. However, as it stands, and with the Combination Problem still looming, it is unclear whether we are sufficiently motivated to adopt Russellian panpsychism over the currently dominant physicalist approach: both have unresolved issues, but the historical explanatory power of the physical sciences might provide inductive evidence that phenomenal experience will come to be explained away through solely physical terms. While this thesis will not provide an explicit resolution to the Combination Problem, in Chapter 3, I shall consider the popular ‘Conceivability Argument,’ which suggests that even with this problem looming, Russellian panpsychism can still provide the most theoretically promising solution to the Mind-Body Problem. While Chapter 4 will find the Conceivability Argument fruitless, Chapter 5 will propose a new and improved argument for Russellian panpsychism, and against physicalism, and conclude that Russellian panpsychism can still offer the best available account of consciousness, even with its Combination Problem unresolved.

2.1 ~ The Traditional Formulation

Russellian monism forms a diverse family of theories united by two theses that both Russell and Eddington defended. In its most fundamental form, there is a positive thesis, and a negative

thesis (Stubenberg, 2016: 58; Goff, 2017: 17-18). The content of these components – taken individually – need not result in Russellian monism; it is through their combination that Russellian monism is born. After explaining these theses and their justifications, I will overview the intended advantage of Russellian monism as a solution to the Mind-Body Problem, and as compared to that provided by physicalism and dualism.

2.1.1 The Negative Component

The negative component of Russellian monism is inspired by Russell and Eddington's claim that physics provides a limited account of the nature of the physical world. Often, this component is presented as the charge that physics only captures the structural, relational, or dispositional properties of fundamental entities, but remains silent as to what those entities are in-and-of-themselves (Eddington, 1920: 200; Alter & Nagasawa, 2015: 425-427; Goff, 2016: 283; 2017: 17).¹² As Russell puts it:

[Physics] lays down certain fundamental equations which enable it to deal with the logical structure of events, while leaving it completely unknown what is the intrinsic character of the events that have the structure... All that physics gives us is certain equations giving abstract properties of their changes. But as to what it is that changes, and what it changes from and to—as to this, physics is silent. (Russell 1959: 18).

Russell proposes here that physics reveals what fundamental entities *do*, but that it remains silent as to what, if anything, these entities *are*. So as not to needlessly complicate matters at this stage, let us adopt a 'bottom-up' perspective of fundamentality, such that micro-level things like electrons are fundamental entities.¹³ Now, let us consider the picture physics paints of an

¹² This claim has also been characterised as the view that physics reveals only the *extrinsic* properties of entities but stays silent as to what *intrinsic* properties they might have (Chalmers, 1996: 153-155; Alter & Nagasawa, 2015: 427). However, such a formulation has broadly fallen out of favour (Ney, 2015; Stoljar, 2015).

¹³ In Section 2.2.3, I will consider how Russellian monism presents when a 'top-down' perspective of fundamentality is assumed.

electron: it is an entity with a negative charge, a certain amount of mass, and can either be free or bound to an atom. Now, let us consider what it is to be a thing with each of these properties: to be a thing with negative charge is to attract things with positive charge and to repel things with negative charge; to be a thing of a certain mass is to resist acceleration and exert attractive forces on other things with mass; and to be free or bound to an atom is to either be or not be in some particular structural relation. In this manner, physics tells us what the role of mass is, but not what the thing is that plays that role; it tells us what the charge role is, but not what the thing is that plays that role; and so on. As Eddington puts it, by the language of physics the electron is – like everything else in physics – ‘a schedule of pointer readings’ (Eddington, 1927: 258).¹⁴ Here, Eddington encapsulates the idea that physics describes what these entities do but stays wholly silent on the matter of what – if anything – underlies these doings.

So far, this should not be a controversial account of the picture painted by physics. Indeed, Russell and Eddington’s observation here is equivalent to the currently dominant thesis within philosophy of science, ‘structural realism.’ Often proposed to be the most defensible form of scientific realism, structural realism is the view that the basic properties described by physics are solely structural/relational/dispositional (Worrall, 1989: 117; Ladyman, 1998). Where the debate begins, is on the matter of whether there is something more to be revealed than that which is provided by this picture. For example, while ‘epistemic’ structural realists propose that science only tells us about the structure of the world (and not about its nature), ‘ontic’ structural realists propose that all that actually exists is that structure (Ladyman, 1998; Ladyman & Ross, 2007; French, 2013). In this manner, while the former view permits that there could be things underlying the structure and relations described by physics, the latter propose that structure is the fundamental nature of reality itself.

¹⁴ I use ‘electron’ henceforth as representative of any truly micro-level entity as posited in the language of physics, whatever that entity (or entities) may come to be.

The Russellian monist endorses epistemic structural realism but combines this with realism about the relevant intrinsic properties, viz. those that ground the physical structure, relations, or dispositions that physics describes (Montero, 2015; Alter & Nagasawa, 2015). They propose that while physics may provide us with a correct abstract representation of reality, it cannot tell us anything about what reality actually is, over and above the fact that it exemplifies a certain formal structure. In this manner, and unlike the ontic structural realist, they believe that there must be something that this structure *applies to* or else the formal system arguably reduces into a sort of uninterpreted or empty formalism (Strawson, 2015: 172; Stubenberg, 2015).

So as to provide an account of the nature of the thing that has that structure, the Russellian monist proposes that the entities described by physics must have some inherent property or quality that grounds the properties of physics, such that one can attach the ‘schedule of pointer readings’ to some real background. Second, and since the language of physics abstracts away from the concrete reality of these entities, the Russellian monist typically proposes that this account must come in a non-structural, non-relational, or categorical (non-dispositional) form. The differences between these terms are not much emphasised in the literature, and many are often used interchangeably (Alter & Nagasawa, 2015: 427). For present purposes, we can say that the Russellian monist believes that we must furnish our account of these entities with some form of *intrinsic nature*, and that this should entail the addition of properties of ‘being,’ not just of ‘doing,’ in such a manner that one may distinguish their bearer from empty space (Schroer, 2010).

It might be intuitive to the reader that, for example, an electron must have intrinsic nature that underlies its behavioural dispositions and relational structure, or, to appropriate Stephen

Hawking, there must be something which ‘breathes fire into the equations’ concerning it (Hawking, 1992: 174). On the other hand, one might believe that there is in fact nothing more to be revealed than the account provided by physics, and so deny the need for such underlying properties, and vice versa. For example, and similar to the views of ontic structural realism, one might be a ‘causal structuralist,’ who argues that the essential nature of an electron and its properties is given by its causal or dispositional profile – viz, by how it disposes its bearers to affect and be affected – and that this is all there is, and indeed, all that needs to be said (Mumford, 1998: 145; 2004; Ellis, 2001; 2002; Molnar, 2003; Bird, 2007a). In this sense, for our electron to be of a certain mass *just is* for it to be disposed to resist acceleration and to attract other masses, such that fundamental properties are reduced to pure causal powers. On this interpretation, fundamental entities are not so much *beings* as *doings*, and so one denies the need for their possession of an underlying intrinsic nature (Goff, 2017: 137).

In order to help motivate the negative component of Russellian monism, one could opt to question how the causal structuralist intends to truly capture the nature of any causal power without referring to properties that are not causal powers themselves (Goff, 2017: 137-142; Campbell, 1976; Robinson, 1982; Blackburn, 1990; Armstrong, 1997; Heil, 2003; Lowe, 2006). We can say that some object *x* has disposition (causal power) *D* to produce outcome *O* in conditions *C*, if and only if, were *C* to obtain, *x* would manifest *O* (Lewis, 1997: 143). In this manner, the nature of a causal power or disposition is defined in reference to that which it gives rise to: when one defines the causal power of, for example, fragility, one does so with reference to the manifestation of something’s breaking; when one defines the causal power of, for example, flammability, one does so by reference to the manifestation of something’s catching fire (Goff, 2017: 137). However, by virtue of its being manifested, that which a causal power gives rise to arguably *is* a categorical property. However, since

categorical properties are properties of being, not of doing, they intuitively cannot be explained solely in terms of causal description. Now, if the nature of that which a causal power gives rise to cannot be captured in this way, *ipso facto*, one cannot fully describe the nature of the power which manifests it likewise; causal structuralism would be mistaken.

To illustrate this, let us consider an object with the causal power of fragility. Now, when manifested, fragility results in that object breaking, and this potential helps define the notion of the causal power. Now, the actuality of the object breaking can be partially defined in terms of further causal descriptions, viz. in terms of the further causal dispositions being broken entails. However, being broken arguably cannot be *wholly* defined in those dispositional terms, as being broken arguably entails a categorical notion of a thing whose being has been impacted upon by those doings: there is a thing which is now broken. Crucially, this categorical property of being is left out of the description provided via causal dispositions. Since that which the power gives rise to helps define the notion of that power, if one cannot fully capture the nature of that which the power gives rise to via causal description, then one cannot also fully describe the nature of the power which manifests it likewise. In this manner, we can see that those who intend to capture the nature of a causal power solely via causal description might leave something crucial out of their description.

For these reasons, there is a sense in which causal or dispositional properties might simply not have enough reality – or at least ontological content – on their own to be all that there is (Bird, 2007b). In the words of Howard Robinson, “a world of powers [is] too insubstantial to command belief” (Robinson, 1982: 114). Similarly, David Armstrong proposes that given a purely dispositionalist account of properties, “particulars would seem to be always re-packing their bags as they change their properties, yet never taking a journey from potency to act”

(Armstrong, 1997: 80). Here, Armstrong suggests that if everything is just causal powers, there is never enough ‘actuality’ in the system.

The causal structuralist might respond that the manifestation of a causal power does not entail some categorical property but instead just entails further causal powers. However, the worry here is that without at some point referring to properties which are not themselves causal powers, one might never reach an adequate description of the original causal power (and hence the fundamental property to be defined): causal power x would be defined by causal power y, which would be defined by causal power z, and so on, ad infinitum, such that we would never get an adequate description of x. Instead, the causal structuralist might then defend that the nature of a causal power is given by its *pattern* of causal relationships, whereby all that it is for something to be a causal power is for it to bear certain relationships to other causal powers (Hawthorne, 2001: 361-378; Bird, 2007b: 513-534). However, while such patterns of relation surely obtain, they again tell us nothing real about the nature of the power in question: by their own language they can still say nothing about what it gives rise to in the real world, as any real explanation of that power’s manifestation is deferred indefinitely. In this sense, the Russellian monist can propose that causal structuralism either involves a vicious regress, or vicious circularity, and so fails to give a satisfactory account of the initial causal power, and by consequence, what that causal power actually is (Goff, 2017: 139-140).

Following such reasoning, and at risk of rendering the concrete world around us unintelligible, the Russellian monist concludes that the entities of physics must then be implemented in something with a nature that goes beyond structural, relational and dispositional description. As Russell puts it:

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. (Russell, 1927a: 325).

In reaching this conclusion, the Russellian monist grounds the negative component of her doctrine: the fundamental entities of physics must be grounded with a nature which *realises* their structures, relations or dispositions. Following custom, let us henceforth call the required property a quiddity:

Quiddities_{ar}: the fundamental, non-relational, non-structural, categorical properties that play the fundamental roles specified in physics.

(Chalmers, 2016a: 26)

To posit quiddities is not to say that an electron has its particular charge, mass, and boundedness *alongside* quiddities, but rather that an electron has quiddities, and that these ground and play the charge role, the mass role, and so on. As argued, without quiddities, one’s worldview appears to reduce into inexplicable bits of ‘empty’ matter, hurrying about in motions that correlate to inexplicable laws expressible solely in terms of their structural relations to one another (Whitehead, 1929b: 49-50; 1938: 131ff). As documented, such a worldview might never be able to give a satisfactory account of what those ‘doings’ actually are, nor what they give rise to, and so might never actually give a satisfactory account of what it intended to.

2.1.2 The Positive Component

The positive component of Russellian monism is its provision of a candidate for quiddities, and which can also allegedly explain how to fit consciousness into our understanding of the physical world (Alter and Nagasawa, 2015: 422). Both Russell and Eddington provide similar candidates for this role, and for the similar reasons. An exact definition, and its form of

realisation at the fundamental level vary amongst the different varieties of Russellian monism. This will be the topic of Section 2.2. For the moment, let us adopt Russell's representative outline that this single 'kind of stuff' "is called mental in one arrangement, physical in the other" (Russell, 1984: 15). In this manner, let us characterise quiddities as properties which either are an instance of or are in some way relevant to conscious experience, and which play particular physical roles (Chalmers, 2016a: 26; Russell, 1927a: 402).¹⁵

The Russellian monist's reason for positing this sort of candidate is twofold. First, they recognise that living brains have an intrinsic nature which involves consciousness, and that such properties of conscious experience are the only class of non-relational, non-structural, categorical property with which one has any direct familiarity. (Russell, 1927a: 402) As Eddington puts it:

We have dismissed all preconception as to the background of our pointer readings, and for the most part we can discover nothing as to its nature. But in one case – namely, for the pointer readings of my own brain – I have an insight which is not limited to the evidence of the pointer readings. That insight shows that they are attached to a background of consciousness (Eddington, 1927: 258-259).

For the Russellian monist, this 'background of consciousness' is the only available candidate of which we have a truly positive conception. Thus, conscious experience – or something very similar – is the best available candidate to fill the role of the intrinsic nature of fundamental entities. Arguably, the only intuitive alternative is a negative one which accepts their intrinsic nature to be incomprehensible. However, to adopt this latter position is to uphold that although our minds have one particular intrinsic nature, the matter that provides their structure has a wholly distinct type of intrinsic nature that is unavailable to our comprehension, and this

¹⁵ How quiddities relate to conscious experience will be explored in Section 2.2, when I delve into the different variants of Russellian monism.

appears to needlessly multiply the number of basic types (Roelofs, 2019: 17; Goff, 2017: 170). Not only then does there seem to be *no good reason* for quiddities not to be received as a serious candidate for the intrinsic nature of fundamental entities, per Occam's razor, they appear to also be the best available candidate for that role.

This brings us to our second reason for positing quiddities: we have two kinds of 'thing' in this world which do not seem to fit together via our current scientific model – mind and matter. Now, as expressed in the previous chapter's dialectic about the Mind-Body Problem and physicalism's Hard Problem, the qualitative phenomenal nature of conscious experience resists all forms of current quantitative physical explanation. Yet, we know conscious experience exists – it is the thing with which we are most closely acquainted – and so it must somehow be fitted back into our understanding of the natural world. Along these lines, it would seem rather foolish to prefer to attach the structure of physics to "something of a so-called 'concrete' nature inconsistent with [mentality], and then to [sit back and continue to] wonder where [consciousness] comes from" (Eddington, 1927: 258-259).¹⁶ Thus, the Russellian monist finalises her synergistic position: quiddities are the non-relational, non-structural, categorical properties of fundamental entities, and quiddities explain consciousness (Goff, 2017: 17-18; Seager, 2006; Strawson, 2008; Coleman, 2009).

2.1.3 The Midground Between our Traditional Theses

Russellian monism is argued to provide an attractive midground between the traditional theses discussed in the previous chapter: it is proposed to combine the virtues of physicalism and dualism, whilst also avoiding their main theoretical vices (Brüntrup & Jaskolla, 2016: 3; Goff, 2017: 18; Alter & Nagasawa, 2015: 422). Now, as noted, Russellian monism forms a spectrum

¹⁶ More shall be said on this in Section 2.1.3

of variants, and some of these are more closely related to dualism over physicalism, and vice versa. In this manner, the strength of some variants to resolve the theoretical issues of physicalism and dualism whilst retaining their respective virtues naturally differ. However, since all Russellian monist theories occupy the theoretical midground between these theses, for the moment, I will consider this midground and its virtues broadly. Granting these potential theoretical virtues, Russellian monism promises to provide an exceptionally attractive position.

It will be helpful here to consider David Chalmers' 'Hegelian-type' argument about this matter, where physicalism forms our *thesis*, dualism forms our *antithesis*, and Russellian monism forms our *synthesis* (Chalmers, 2016a: 20). As noted in Chapter 1, physicalism has the virtue of integrating consciousness into the natural order and is capable of explaining many kinds of mental phenomena. However, it suffers a Hard Problem whereby it appears to simply fail to explain the existence of consciousness itself. If one recalls, I delineated the tasks of mechanistically explaining mental phenomena as the 'easy problems,' and the task of accounting for the nature of conscious experience as the 'Hard Problem' (Chalmers, 1995: 200-219). We saw that physicalism has the virtue of being capable of resolving the easy problems of consciousness, but the vice of the Hard Problem. Now, dualism has the virtue of avoiding the Hard Problem: by positing consciousness as something fundamental in-of-itself, it does not distort or disregard its nature when attempting to explain it. However, it has the converse vice of separating consciousness from the natural order and so is potentially *incapable* of integrating consciousness within physical causation and of resolving issues of over determination. In this manner, we can say that the argument for the thesis are the causal arguments against dualism, and the argument for the antithesis is the Hard Problem against physicalism (Chalmers, 2016a: 20).

The Russellian monist agrees with the physicalist that consciousness is part of the physical world. In this manner, she appears to avoid the dualist's vice of separating consciousness from the natural order and thus from problems of causation. However, she also agrees with the dualist that the dispositional properties described by current physical science are themselves insufficient to account for consciousness. By adding quiddities, she grounds consciousness at the fundamental level, and so appears to avoid the physicalist's Hard Problem of consciousness, viz. of explaining *how it is* that things are conscious (Lee, forthcoming: 13; Cutter, 2019: 113; Goff & Coleman, 2020: 304-305). As noted, quiddities are the underlying properties of fundamental entities: physics as we know it is grounded in their relations, and consciousness, as we know it, is intimately related to their intrinsic nature. Quiddities – and so consciousness – are then integrated into physical causation, and since they instantiate or are in some way relevant to consciousness, they embody and so preserve its distinctive features.

We may now yield our 'synthesised' argument for Russellian monism: it captures the virtues of our thesis and antithesis, whilst avoiding their respective vices (Chalmers, 2016a: 20). I will treat this as a major potential advantage over our traditional theses. Indeed, the Mind-Body debate persists because of the respective vices of our traditional counterparts. Any theory which could harness each of their virtues, whilst avoiding their vices should necessarily then be treated with priority. Accordingly, I continue under the assumption that if Russellian monism can achieve this advantage without suffering some unresolvable and major theoretical vice, it would provide a more satisfactory account of consciousness than its traditional counterparts, complete with insights which might genuinely disband the Mind-Body Problem (Holman, 2008: 49).

2.2 ~ The Variants of Russellian Monism

Russellian monism forms a spectrum of variants, each of which upholds the positive and negative component at its core. Broadly speaking, there are three main dimensions of difference amongst its contemporary forms: (i) panpsychism versus panprotopsyism; (ii) constitutive versus emergent; and (iii) micropsychism vs cosmopsychism. The stances in (i) distinguish the particular form of intrinsic nature quiddities instantiate; the stances in (ii) distinguish the fundamentality of our conscious experience alongside its relation to fundamental forms; and the stances in (iii) consider broader ontological questions about fundamental entities and directions of ontological dependence. Each of these dimensions are orthogonal to each other, and so stances within one dimension can be paired with stances from other distinct dimensions, barring a few exceptions. For example, one could be a constitutive *or* emergentist Russellian monist, whilst also being a panpsychist *or* panprotopsyist. I will introduce each of these dimensions in turn and clarify at each stage onto the stances which I believe to be most theoretically promising. Each adopted stance will then be carried on into my explanation of the subsequent dimension(s), until we finalise a precise definition of the Russellian monism of interest to this thesis: a constitutive and micropsychist form of Russellian panpsychism.

It will be helpful to quickly reemphasise here that we are still only interested here in *phenomenal experience*, as opposed to the more sophisticated phenomenon of *conceptual thought*. Where the former refers just to the ‘something it is like’ to be a conscious thing, the latter refers to any of the particular psychological attitudes towards specific propositions one enjoys as a higher-order conscious thing. When adopting the former focus as our understanding of consciousness, the Russellian monist calls this focus *panexperientialism*, and presents it as the thesis that experience – or something close to it – is fundamental and ubiquitous. When adopting the latter focus, the Russellian monist calls this focus *pancognitivism*, and presents it

as the thesis that thought – or something close to it – is fundamental and ubiquitous. Now, and as noted in Chapter 1, I believe that the ‘something it is like to be’ a conscious thing is the most crucial aspect to be accounted for in our explanation of consciousness. Once you have accounted for such a phenomenon and instantiated it within some relevant structure, it might be the case that ‘thought’ could be explained away as an ‘easy problem,’ and so experience would precede thought by logical necessity in our explanation. Regardless of whether this obtains, ‘phenomenal experience’ was proposed to be the logical starting point to any explanation of consciousness. For this reason, all subsequent discussion about Russellian monism refers to the panexperientialist variety, just as my reference to ‘consciousness’ continues to refer to phenomenal experience.

2.2.1 Russellian Panpsychism versus Panprotopsychism

As traditionally construed, panpsychism (from the Greek pan- and psyche, or ‘all-’ and ‘mind’) is the thesis that mind is a fundamental feature of the world. This notion precedes Russellian monism and can be found throughout the history of philosophy; one may be panpsychist without being Russellian (Skrbina, 2005; Brüntrup & Jaskolla, 2006: 1). For example, varieties of panpsychism may be found within the Presocratic texts of Thales (c. 625-545 BCE) (Barnes, 1982: 96-7), and also the Enlightenment thinkers of Baruch Spinoza (1632-1677), Gottfried Wilhelm Leibniz (1646-1716), and, as I have recently argued, Anne Conway (1631-1679) (Lee, forthcoming). For the most part, however, the thesis reached its prominence during the nineteenth century with the work of Gustav Fechner (1801–1887), Wilhelm Wundt (1832–1920), Rudolf Hermann Lotze (1817–1881), William James (1842–1910), Josiah Royce (1855–1916) and William Clifford (1845–1879). One could argue that we are now living in the beginning of another revival. The work of Philip Goff (2017; 2019), David Chalmers (1996;

2010), Galen Strawson (1986; 1994), and Luke Roelofs (2019) – to name but a few – have been instrumental in establishing its reintroduction into the philosophy of mind debate.

Contemporary panpsychism is the theory that phenomenal experience itself is fundamental and ubiquitous. This does not necessarily entail that inanimate objects such as plates, or geographical areas such as County Durham experience in the relevant sense. However, it does entail that they are grounded by fundamental entities which do. Continuing to assume that things like electrons are fundamental, our electron *would* be posited to have some exceptionally rudimentary form of experience that is radically different in complexity to that with which we are acquainted (Strawson, 2008; 2015). Furthermore, contemporary panpsychism does not entail that conscious experience is the *only* fundamental and ubiquitous feature of the universe, and so should be distinguished from more traditionally idealistic notions where just consciousness is the ground of all reality (Brüntrup & Jaskolla, 2016: 1).

Building from this, the Russellian panpsychist posits quiddities to be a form of phenomenal property. As opposed to the ‘macro-’phenomenal properties we find in our macro-experience – viz. our usual conscious experience – she believes that these underlying phenomenal properties are a form of ‘micro-’experience – properties characterising what it is like to be a micro-physical entity.¹⁷ Per Russellian panpsychism, quiddities are then micro-experiential determinates of the determinable property of consciousness (Cutter, 2019: 114). They ground and play the roles associated with fundamental entities, and also instantiate the phenomenal properties that feature in our macro-experience. We can formulate a precise definition as follows:

¹⁷ As noted, and so as not to overcomplicate matters, we continue here to adopt a ‘bottom-up’ perspective of fundamentality. However, so as to remain accurate, each formal definition will be construed in terms that are open to be interpreted from either view. The final dimension will then introduce a different ‘top-down’ perspective of fundamentality.

Russellian panpsychism: phenomenal properties underlie and characterise the fundamental constituents of reality, and from which, conscious experience can be somehow derived.¹⁸

Panprotopsychism, on the other hand, was coined by David Chalmers in his discussion of quiddities in *The Conscious Mind* (1996). Like panpsychism, one can be panprotopsychist without being Russellian. By itself, the core theory upholds that conscious experience is not itself fundamental and ubiquitous, but that something intimately related to it is instead. Building from this notion, Russellian panprotopsychists take quiddities to be ‘protophenomenal.’ By protophenomenal, they refer to properties which are not themselves phenomenal, but which hold some exceptionally close relationship to phenomenal properties, such that they are related to conscious experience in the same way that basic physical ‘role’ properties are related to non-basic properties such as temperature (ibid: 126-127).¹⁹

One might be immediately concerned that Russellian panprotopsychism is simply then a form of physicalism. Indeed, both type-b and -c physicalism entail that fundamental properties are not phenomenal properties, but that they ground phenomenal properties when appropriately arranged. However, the thought behind Russellian panprotopsychism is that protophenomenal properties are a special new form of property, which hold an especially close connection to phenomenal properties. In this manner, we can say that (i) protophenomenal properties are distinct from structural properties, and (ii), that there is an a priori entailment from truths about

¹⁸ I leave the notion of ‘derive’ open here, as it shall be the focus of the next dimensions.

¹⁹ There is also a close cousin of panprotopsychism called ‘panqualityism.’ Introduced by Herbert Feigl (1960: 24-36), and stemming from his conversation with Stephen C. Pepper, it has been most thoroughly developed and endorsed by Sam Coleman (2015, 2016). For Coleman, it is ‘qualities’ that are fundamental. Although Coleman’s work warrants a much more robust explanation, due to the particular focus of my thesis, I shall refrain here.

protophenomenal properties (with, perhaps structural properties also) to truths about the phenomenal properties they constitute (Chalmers, 2016a: 31). By ‘a priori entailment,’ Chalmers refers to a logical relationship whereby the truth of the consequent can in principle be deduced solely through reasoning regarding the antecedent. In this case, Chalmers proposes that truths about phenomenal properties can be deduced solely through reasoning concerning truths about protophenomenal properties (with, perhaps structural properties), at least in principle.

Russellian panprotopsychism does not entail that things like electrons have any particular phenomenal experience; there is nothing it is like to be a single protophenomenal property (Chalmers, 2016a: 31). Instead, one achieves facts about phenomenal properties from facts about protophenomenal properties, structural properties, and particular appropriate conditions. In this manner, protophenomenal properties ground and play the roles associated with fundamental entities, and whilst not themselves phenomenal, can come to instantiate forms of experience. We can formulate a definition of Russellian panprotopsychism as follows:

Russellian panprotopsychism: non-structural protophenomenal properties underlie and characterise the fundamental constituents of reality, and are intrinsically suited to bringing about phenomenal properties, when under the appropriate conditions.

In this thesis, I shall adopt and argue for Russellian panpsychism, as opposed to Russellian panprotopsychism. This is not to say that the latter is not an attractive theory of consciousness, nor that it might not turn out to be the correct one. Rather, I simply believe that the case for the panpsychist variant of Russellian monism is more analogous with our initial motivation for Russellian monism and the advantages associated with it. In Section 2.1.2, I discussed the

necessity of a candidate for the intrinsic nature of fundamental entities. I presented that a major virtue of Russellian monism here, is its provision of a *positive* candidate, which is both accessible to our introspection – and so positively conceivable – and which also maximises theoretical simplicity by not needlessly multiplying the number of basic types. Following such argumentation, and if Russellian monism is to provide a theoretically more promising solution to the Mind-Body Problem than physicalism, I believe that phenomenal properties are a better initial candidate for the intrinsic nature of fundamental entities than protophenomenal ones.

My reasoning stems from the obscurity that arises from any attempt to define ‘protophenomenal.’ As noted, we are directly familiar with phenomenal properties. We do not, however, have a positive conception of protophenomenal properties, as – and as charged to physicalism – we cannot conceive of what it is like to be something that brings about experience whilst not being experiential itself (Chalmers, 1996: 154). However, we are supposed to accept that although this notion may be inaccessible to us, if we did understand it, we could understand how conscious experience could arise from it under certain conditions. Now, the panprotopsychist could present that protophenomenal properties are only currently inconceivable, and so that we might at some point come to form a positive conception of them (Pereboom, 2011: 117-118). Or they could propose that they are simply inconceivable and so will never be accessible to our introspection (McClelland, 2013). Now, while the former option might hold some fruit to the Russellian monist, it remains unclear how and if such a conception could ever come about. Furthermore, it diminishes the theoretical simplicity of the theory over panpsychism by positing an additional unfamiliar basic type. Similarly, if one adopts the latter claim that such properties are inconceivable, they would necessarily fail to establish both base tenets of the positive dimension of Russellian monism – a positively conceivable candidate for the intrinsic nature of fundamental entities that maximises theoretical simplicity.

While I shall adopt and argue for Russellian panpsychism in this thesis, quiddities as protophenomenal properties might yet bear fruit to overcome the Mind-Body Problem.²⁰ For the purpose of this inquiry, however, and in the spirit of remaining analogous with the foundational components of Russellian monism, I adopt Russellian panpsychism henceforth, and carry it on into my explanation of the next major dimension of difference.

2.2.2 Constitutive versus Emergent Russellian Monism

The distinction between constitutive and emergentist forms of Russellian monism concerns the fundamentality of our macro-experience, and its relation to fundamental entities. Now, for the constitutive Russellian panpsychist, conscious wholes are transparently composed by conscious parts, such that the produced is nothing over and above the producer (Goff, 2017: 150). In this manner, one's macro-experience is not itself fundamental, and should instead be understood in terms of facts about more fundamental kinds of consciousness.

Constitution_{df.} some entity E, and its property instances P, are constituted by entities F1... Fn, and their property instances Q1...Qn, iff:

- (1) Entity E and its property instances P exist in virtue of F1... Fn, and their property instances Q1... Qn, and the real relations R that exist between them.²¹ And,
- (2) Entity E and its property instances P are nothing over and above F1...Fn, their property instances Q1...Qn, and those real relations R.

²⁰ Sam Coleman, for example, (2015; 2016), suggests that protophenomenal properties might be identified with qualities, akin to those we perceive, such as redness, and of which we *do* have a positive conception.

²¹ By 'real relations' I refer to those relations that obtain independently of human cognition. That is, those that exist in the world irrespective of whether they are thought about, conceptualized, or linguistically expressed (Aquinas, 1947: I, q.13, a.7; Leibniz, 1898: 47-50; Armstrong, 1997: 85; Lowe, 1998).

Now, constitution is the most fundamental building relation we understand in the world. For example, a table is typically constituted by four legs, a flat top, and the particular relations existing between those parts so as to instantiate that ‘table-like’ structure; the table leg is constituted by a particular array of molecules and the particular relations they stand in so as to instantiate that ‘table-leg-like’ structure; those molecules are constituted by a particular group of atoms, bounded together in some real relation; and, those atoms are constituted by a particular group of protons and neutrons standing in a particular relation so as to instantiate a nucleus, and a number of bound electrons standing in another set of particular relations so as to surround that nucleus. Our table is nothing over and above those entities (and their property instances) which constitute it. In this manner, one may find examples of constitution wherever, and however they look.

The constitutive Russellian panpsychist extends this form of building relation to our conscious experience, such that it is not itself fundamental, but exists in virtue of, and as nothing over and above, more fundamental forms of conscious experience. Throughout this exposition, we have adopted the most commonly assumed direction of constitution and fundamentality – one which explains larger non-fundamental things to exist in virtue of, and as nothing over and above smaller fundamental things (such that things like electrons go on to constitute things like tables). However, the constitutive Russellian panpsychist could also uphold a converse view, such that smaller non-fundamental things exist in virtue of, and as nothing over and above larger fundamental things. As noted, the particular direction this relation of fundamentality takes will be the focus of the next subsection. In this manner, and for the moment, we can formulate the following broad definition:

Constitutive Russellian panpsychism: phenomenal properties underlie and characterise the fundamental constituents of reality, and ground and constitute our macro-experience such that it exists in virtue of, and as nothing over and above them.

Emergentist Russellian panpsychism understands macro-experience to emerge from fundamental forms of experience. Usually, the direction of dependence is construed in terms of smaller fundamental things (such as quiddities) sustaining larger fundamental things, as opposed to the converse, and so the emergentist is not usually committed to discussions about smaller things emerging from and being sustained by larger things. The typical emergentist Russellian panpsychist, therefore, proposes that macro-experience had by complex physical entities emerges from the micro-experience of its parts, in accordance with certain emergence laws that directly connect complex consciousness with the simpler kinds instantiated by micro-level entities (Roelofs, 2019: 18). While the most basic physical entities are then still micro-experiential, the macro-experience had by complex physical entities cannot be defined solely in terms concerning the constitution of their parts.

Now, emergence is typically understood to come in two forms. The first, and the most commonly discussed within science, is weak, or ‘epistemic’ emergence. We can say that something is weakly emergent when the high-level phenomenon arises from the low-level domain, but that truths concerning that phenomenon are *unexpected* given the principles known to govern the low-level domain. Nevertheless, they are also nothing over-and-above and are in principle a priori entailed by them. A useful example here is the emergence of high-level patterns – such as translating oscillators, or ‘gliders’ – in cellular automata: if one is only provided with the low-level rules governing the automaton, these high-level patterns are unexpected; however, following advanced calculation, they can be seen to be entirely deducible

from, and a consequence of the low-level facts (Chalmers, 2006: 244-246; Bays, 2009). While weak emergence is an interesting phenomenon then, an omnipotent metaphysician would see it as nothing over the constitution principle expounded above; upon ideal reflection, weak emergence *just is* constitution. For this reason, I omit it from further consideration and distinction.

The second type of emergence, and that of interest here, is strong or ‘ontological’ emergence. As opposed to the weaker form, this form of emergence is that which is usually referred to when invoked within philosophical discussion. We can say that something is strongly emergent when the high-level phenomenon arises from the low-level domain, but that truths concerning that phenomenon are *not deducible* even in principle from truths about the low-level domain (Chalmers, 2006: 224-246). Strong emergence requires that high-level truths are not metaphysically necessitated by low-level truths. Unlike weak emergence, however, there are no agreed upon or clear-cut examples of this within the natural world. Indeed, the discovery of a genuine instance would require an expansion to our conception of the physical world, as it would necessitate fundamental laws not already posited or predicted. However, in light of the enduring Mind-Body debate and the particular nature of conscious experience, the phenomenon of consciousness is perhaps the most plausible available candidate for this form of emergence.

Strong Emergence_{ar}: some object or property instance P is strongly emergent from some entity E, with property instances Q1... Qn iff:

- (1) P is an object or property instance which is causally produced or sustained by E in accordance with certain fundamental emergence laws, and,
- (2) P is something over and above Q1... Qn, and the real relations R between them, or,

(3) P cannot be derived from Q1... Qn, and the real relations R between them when so combined.²²

A strongly emergent property is then ontologically novel and so can be considered to be fundamental in-of-itself. As before, the direction of emergence is typically construed in terms of smaller fundamental things sustaining larger fundamental things, as opposed to the converse, and so the strong emergentist is not usually committed to discussions about smaller things emerging from and being sustained by larger things.²³ Instead, she typically understands that there are just small fundamental things, and larger fundamental things, the former of which sustain the latter in accordance with certain fundamental emergence laws. For the typical emergentist Russellian panpsychist then, quiddities sustain macro-experience, whilst macro-experience is not wholly grounded in, nor a priori deducible from quiddities. Now, there are ‘layered’ views of this emergentism, where macro-experience is posited to co-exist with the relevant quiddities (Rosenberg, 2004; Brüntrup, 2016), and ‘fusionist’ accounts, which posit that the micro-experience of our micro-level parts fuse to become our macro-experience and cease to exist in the process (Seager, 2016; Mørch, 2014).²⁴ For the purposes of this discussion, the following broad definition of emergentist Russellian panpsychism will suffice:

Emergentist Russellian panpsychism: phenomenal properties underlie and characterise the fundamental constituents of reality, and when appropriately governed

²² As noted by Goff, there might intuitively be cases of ‘intelligible emergence,’ which would satisfy (1) and (2), but not (3). Intelligible emergence is where a fundamental lower-level property intelligibly causes and sustains a higher-level property in such a way that one could move a priori from the instantiation of the former to the instantiation of the latter (Goff, 2017: 150-151). In this instance, one could instance strong emergence without necessitating (3). Nevertheless, (3) necessarily entails (2), such that we can say that either (1) and (2), or (1) and (3) are necessary for strong emergence.

²³ Although such a stance is perhaps metaphysically peculiar, it is theoretically available in principle.

²⁴ Mørch’s view differs slightly from Seager’s in as much as Seager views the conscious brain to entirely lack parts, whereas Mørch simply believes the parts of post-fusion brain are dependent for existence upon the brain.

by certain fundamental emergence laws, give rise to ontologically distinct conscious experience without any a priori entailment.

In this thesis, I shall adopt and argue for constitutive Russellian panpsychism as opposed to the emergentist variety. Again, this is not to say that the latter might not turn out to be valid. Rather, I simply believe that the constitutive variety has the potential to offer a greater explanatory advantage over our traditional theses. My reasoning is twofold. First, by positing macro-experience as ontologically novel, our emergentist Russellian panpsychist appears to add a similar explanatory gap to the physicalist, and so potentially lessens her claim against him. Russellian monism promised to account for the distinctive nature of our conscious experience and proposed this to be a robust theoretical advantage via the charge that the physicalist cannot. However, if micro-experience does not itself ground or transparently explain macro-experience, one might question why its explanation about that macro-experience should be taken as superior; it simply posits an additional fundamental property type which still does not itself transparently explain the phenomena in question. Furthermore, and as noted in Chapter 1, I do not entertain emergentist physicalism in this thesis. In principle, however, our physicalist could also adopt emergentism and then further question the theoretical advantage emergentist Russellian panpsychism actually provides: if quiddities do not transparently explain macro-experience, why not just maintain that macro-experience comes from nonconscious microphysics in an equally puzzling way, and without added metaphysical commitment?

Now, I do believe that emergentist Russellian panpsychism would still fare better than physicalism with regards to this issue. In the emergentist physicalist's case, experiential phenomena are grounded in, or – following the former paragraph – perhaps emerge out of, nonexperiential phenomena, in a way that appears to be as extraordinary (and unlikely) as the

instantiation of spatial phenomena by nonspatial phenomena (Strawson, 2008: 64). By positing quiddities, however, the emergentist Russellian panpsychist makes it the case that there is something about entity E and its property instances $Q_1 \dots Q_n$, such that although P is something over and above E and its property instances, the property that emerges (macro-experience) is intimately related to the property instances instantiated by E (micro-experiences); its emergence is not an outright miracle. In this manner, I do not believe it should be discounted as a viable option over the physicalist alternatives. However, since constitutive Russellian panpsychism does not have an explanatory gap of this kind at all, *ceteris paribus*, it has the potential to offer a larger explanatory advantage than emergentist Russellian panpsychism.

The second part of my reasoning for adopting constitutive Russellian panpsychism concerns the fact that the emergentist variety might face a similar problem of causal exclusion to dualism. If one recalls, this is the issue of reconciling causal closure at the micro-level with the causal efficacy of macro-experience. Consider some material event E caused by a macro-experiential fact (M_a), for example, my gasping due to the feeling of pain after stubbing my toe. Given causal closure, this event must have a sufficient micro-level cause (M_i). Now, for the constitutive Russellian monist, one's macro-experience is nothing over and above facts about material parts and quiddities. Accordingly, overdetermination in the case above is superficial, since M_a is nothing over and above M_i , and so E is not somehow caused twice over (Goff, 2017: 156; Roelofs, 2019: 19). However, if one's macro-experience is ontologically novel, as per emergentist Russellian panpsychism, it is fundamental in-of-itself: it either co-exists with the fundamental quiddities which sustain it, or it replaces them upon emergence. In the case of the former, M_a co-exists with M_i , and so both sufficiently cause E, and so E is overdetermined. In the case of the latter, M_a replaces M_i , and so while safe from

overdetermination, E does not achieve causal closure at the micro-level.²⁵ In either sense, it appears that emergentist Russellian panpsychism might fail to resolve the problem of causal exclusion. *ceteris paribus*, constitutive Russellian panpsychism appears to provide the more attractive, and potentially more viable midground.

2.2.3 Micropsychism versus Cosmopsychism

Our final distinction concerns the metaphysical ‘placement’ of fundamental entities and the direction of ontological dependence. There are two plausible candidates within this discussion. We should first assume that our universe has parts, and a structure of dependence relations – the general assumption of most metaphysicians (Miller, 2018: 5). It follows that the behaviour of these parts, the dependence relations between them, and any event about them, can be theoretically explained in terms of universal laws since they are the totality of that which the universe consists of. Now, since these laws are – or at least would be universal – they must be grounded in the causal powers of something truly fundamental, and intuitively, such an entity or entities must be at one end of the spectrum from the littlest things (absolute micro) to the largest things (absolute macro) (Goff: 2017: 248). Along these lines, the two most plausible candidates for our position of fundamentality are either the totality of entities, viz. the cosmos itself, or the smallest micro-level entities about that totality.

Traditionally, and as noted previously, the common assumption of philosophers and scientists has been the latter: that fundamental things exist at the micro-level (Coleman, 2006).²⁶ Indeed, much of the previous exposition about Russellian monism has been formulated around such an

²⁵ A similar case could be made for the ‘top-down’ emergentist, whereby our consciousness emerges out of a larger fundamental consciousness, and so results in overdetermination again.

²⁶ It must be noted, however, that contemporary theoretical physics posits wave function and fields as fundamental, and so might draw some greater, or at least partial analogy to the former assumption: that fundamental things exist at the universal level.

assumption.²⁷ Formally construed, this assumption amounts to the thesis that the micro-level entities which form proper parts of the universe are the fundamental entities of our universe, and that facts about them ground facts about macro-level entities. The table analogy in Section 2.2.2 serves as a good representation of this fundamentality, and the subsequent direction of dependence about it: fundamental micro-level entities ground facts about atoms, facts about molecules, and eventually, at some abstract level of complexity, facts about table-legs and tables. In this manner, we can say that all things exist and are the way they are in virtue of how those micro-level entities are (Goff, 2017: 19). When extended into Russellian monism, this thesis becomes ‘micropsychism:’ or, the thesis that micro-level quiddities are fundamental. Applied to constitutive Russellian panpsychism, this entails that macro-experience exists in virtue of, and as nothing over and above the phenomenal experience of quiddities; facts about our macro-experience are nothing over and above facts about the phenomenal properties of the micro-level entities which constitute our brains and bodies. We can formulate a definition of a constitutive and micropsychist form of Russellian panpsychism as follows:

Constitutive micropsychist Russellian panpsychism: phenomenal properties underlie and characterise the fundamental entities posited by physics in its characterisation of matter, and ground and constitute macro-experience in such a manner that it exists in virtue of, and as nothing over and above them.

The second candidate for the placement of fundamentality is at the cosmos-level. Heavily influenced by the work of Jonathan Schaffer (2010), this view upholds that everything within the universe – such as tables, table legs, molecules, atoms, and so on – exist as they are in virtue of the universe existing as it is. A useful analogy here is to consider an omnipotent being

²⁷ This thesis underlying this assumption is sometimes referred to as ‘smallism’ (Coleman, 2006) or as ‘priority pluralism’ (Miller, 2018).

creating the universe as a singular and concrete totality in such a way that everything within that totality exists as it does only because of the way that totality is (Goff, 2017: 233). Here, the totality itself would be fundamental; while smaller things would still compose larger things, those smaller things would exist in virtue of that which they compose at the universal scale. To help illustrate this better, we can modify our grounding relation here to now refer to ‘*grounding by subsumption*’ (Goff, 2017: 234).²⁸

We can say that some entity X *grounds by subsumption* entity Y iff: (1) X grounds Y, (2) X is a unity of which Y is an aspect, and (3) Y is nothing over and above X (Goff, 2017: 221). Consider the specific shade of orange – ‘orange 7 (O7)’ – which involves red hue (R), yellow hue (Y), a certain degree of saturation (S_x) and some certain degree of lightness (L_x). One could suppose via grounding by analysis that this colour is a composite property made up of such hues, and degrees of lightness and saturation. Conversely, one could suppose via grounding by subsumption that O7 is a unified property of which those constituents are aspects, whereby:

- O7 grounds R, Y, S_x, and L_x
- O7 is a unified property of which R, Y, S_x, and L_x are aspects (ibid)
- R, Y, S_x, and L_x, are nothing over and above O7.

For present purposes, we can say that per priority monism, all entities are grounded by subsumption in facts about a singular and universal macro-level entity. In this manner, the universe is a fundamental unified whole, and all other material entities are aspects of, and

²⁸ Unless otherwise, ‘grounding’ will still continue to refer to ‘grounding by analysis’ for the remainder of this thesis.

nothing over and above that whole. While Schaffer calls this thesis ‘priority monism,’ when combined with panpsychism, the combined thesis is often called ‘cosmopsychism.’²⁹

When extended to constitutive Russellian panpsychism, cosmopsychism amounts to the thesis that our macro-experience exists in virtue of, and as derivative of a fundamental and universal consciousness. In this manner, facts about consciousness are grounded by subsumption by consciousness-involving facts concerning the universe (Lee, forthcoming: 18). Now, this is not to say that the universe is *a thinking conscious thing* – our focus remains solely on the panexperientialist variety of Russellian monism. Rather, it is just to say that the universe is a fundamental entity that has experience of which our macro-experience is a proper part. In this manner, the constitutive Russellian panpsychist could uphold that this universal consciousness is an incoherent lump of micro- and macro-experiences which do not add up to anything like cognition. We can thus formulate a definition of a constitutive and cosmopsychist form of Russellian panpsychism:

Constitutive cosmopsychist Russellian panpsychism: the universe is a fundamental, unified and experiential whole, which grounds macro-experience by subsumption.

In this thesis, I will adopt a micropsychist perspective about constitutive Russellian panpsychism. First and foremost, my reasoning stems from my belief that the micropsychist model might be initially more palatable within general scientific and metaphysical assumptions about fundamentality. As noted, contemporary theoretical physics posits wave function and fields as fundamental, and so might draw some greater, or at least partial analogy to the

²⁹ Recent further defences of this position can be found in Goff (2017), Nagasawa and Wager (2016), and Shani (2015).

assumption that fundamental things exist at the universal level. However, it remains an intuitive assumption to many that little things compose larger things, and that this is the typical direction of dependence regarding fundamentality. Accordingly, if Russellian monism is to be integrated into science as a genuine model for resolving the Mind-Body Problem, a micropsychist bent might initially increase its accessibility.

Again, this is not to say that cosmopsychism might not turn out to be correct. Indeed, the nature of fundamentality likely necessitates that one or the other approach is correct – regardless of whether one applies them to Russellian monism. Notwithstanding its own issues, cosmopsychism arguably provides viable resolutions to many of the issues faced by micropsychism. Furthermore, the cosmopsychist could make a similar argument for constitutive Russellian panpsychism to that which I will level in this thesis. Although I am then somewhat agnostic as to which will turn out to be most promising, I adopt the constitutive and micropsychist form of Russellian panpsychism henceforth since I believe it is the best place to start.

2.3 ~ Considerations and Concerns

For the sake of clarity, the constitutive and micropsychist form of Russellian monism I have adopted shall henceforth be referred to as *Russellian panpsychism*. It will be useful to briefly restate where this position fits in the Mind-Body debate. Clearly, it upholds the positive and negative theses of Russellian monism – providing a positive candidate for the intrinsic nature of micro-level entities and which allegedly fits consciousness into our understanding of the physical world – and so satisfies our initial motivation for the doctrine. It also appears to hold a relatively central midground position between our theses of dualism and physicalism and so appears to enjoy all the benefits associated with that midground. First, it appears to be protected

from physicalism's Hard Problem; by positing quiddities as phenomenal properties, one grounds macro-experience in entities which transparently embody and so preserve the features that the physicalist fails to account for. Second, since we have adopted a strictly constitutive building relation about these quiddities, viz. constitution about parts as opposed to emergentism, Russellian panpsychism also appears to be immune to the dualist's issues of causation and overdetermination.

As previously commented, any theory which can harness the virtues of physicalism and dualism, whilst avoiding their vices should be treated with priority; Russellian panpsychism might provide a genuine resolution to the Mind-Body Problem. However, the theory does have its own problems, and it must resolve these for its potential resolution to obtain. In this section, I shall address two of the most commonly pressed. First, I will consider and reject the worry that Russellian panpsychism is at odds with science. Second, I will consider the infamous 'Combination Problem.' I shall argue that this problem constitutes a genuine threat to the credibility of Russellian panpsychism, which may diminish the motivation to adopt it over the currently dominant physicalist approach.

2.3.1 Is Russellian Panpsychism at Odds with Science?

Upon presentation with Russellian panpsychism, one might be quick to adopt a cautious distrust. One might be suspicious of the claim that physical science has its limitations and find the additional positing of electrons as experiential to have "the faintly sickening odour of something cooked up in the metaphysical laboratory" (Nagel, 1986: 49). As Roelofs acknowledges, history and current affairs rightly teach us that all claims of the form 'science is great, but...' should be taken with suspicion: beliefs of this form hampered our response to the Covid-19 pandemic, are holding back progress at tackling climate change, and indeed,

nearly overthrew US democracy (Roelofs, 2021: 117). In this manner, we should be reluctant to give unscientific-sounding theories the benefit of the doubt. While I am entirely in agreement with this reasoning, I believe that extending this suspicion to Russellian panpsychism is misplaced. I shall argue that Russellian panpsychism is compatible with present science, and the direction with which science is progressing. I will also address and dismiss the worry that it is potentially unfalsifiable.

One way a theory can avoid claims that it is unscientific is through showing its compatibility with science. Now, it is true that Russellian panpsychism is interested in something that science allegedly cannot account for. However, the doctrine is also quick to acknowledge exactly what science *can* tell us, specifically in terms of causal structure, such as what causes what, when, where, and how (ibid: 118). Indeed, its proponent acknowledges and strives for compatibility with all that science proposes and predicts. In this manner, one can distinguish Russellian panpsychism from pseudoscience and scientific conspiracies that are characterised by denials of science. For example, the claim that climate change is not real is a denial of robust scientific data. Similarly, séances and telepathy require causal mechanisms that are at odds with those predicted by science. Russellian panpsychism, on the other hand, treats science as the final authority on all matters of the world's causal structure, and by positing quiddities does not add something which is at odds with that, but rather something that was allegedly there all along; quiddities sit outside of the current method of scientific inquiry, but not outside the fundamental reality of interest to that inquiry. In this manner, Russellian panpsychism is not incompatible in principle with science.

Russellian panpsychism is also striving for compatibility with future scientific advancements. The purpose of science is to understand, and this understanding advances in accordance with

the available data and the suitability of theories to explain such data. Integral to science is then the ability to revise itself about its weaknesses when new phenomena appear that are at odds with its dominant theories, or when its dominant theories are revealed insufficient to fully account for the available data. Russellian panpsychism argues that current science, as a precursor to the physicalist discipline, cannot account for the phenomenal nature of conscious experience. Indeed, and as previously mentioned, all that such methods can tell us about consciousness distort and disregard its distinctive nature such that their explanation is consistent with an absence of the data intended to be understood. Now, many would like to say that this particular explanatory gap will be explained away as science progresses. I would also. However, there is good reason to think it will not be explained away under the current approach. If we are willing to amend our current worldview, future science might explain away this problem, but in light of the continuing Mind-Body Problem, current approaches to science might not. Accordingly, we must allow for our scientific worldview to be amended so as to remedy its inability to account for the data of consciousness, and indeed, so as to be scientific about the available data; Russellian panpsychism simply provides a potential candidate for this revision.³⁰

For further emphasis, it will be useful to acknowledge that physicists have, on occasion, posited something analogous to consciousness at the fundamental level. To name but a few: Roger Penrose and Stuart Hameroff proposed that the collapse of wave function is a moment of experience (Penrose, 1996; Hameroff & Penrose, 1996); Max Planck proposed that consciousness must be fundamental, and that matter is derivative from it, and claimed that “We

³⁰ Goff (2017, 2019) interestingly proposes that the success of science has come through *limiting the scope of our inquiry*, and so that confidence in its provision of universal understanding would be mistaken. Citing Galileo’s push for an exclusively mathematico-nomic scientific worldview, he concludes that current science is not equipped to deal with conscious experience, and that Russellian monism might provide the only way of reintegrating consciousness back into our scientific worldview (Goff, 2017: 11-14; Galileo, 1623: 237-238).

cannot get behind consciousness” (Planck, 1931: 17); Erwin Schrödinger stated that “[T]he material universe and consciousness are made out of the same stuff.” (Schrödinger, 1931: 15-16). In this manner, the notion of consciousness at the fundamental level is at least not alien in principle to physics. Similarly, and although not scientists themselves, Godehard Brüntrup and Ludwig Jaskolla acknowledged that in the nonstandard Bohmian interpretation of quantum mechanics, particles are informed by the ‘active information’ available within quantum potential, which is ultimately capable of representing information as within the limits provided by Schrödinger’s equation, and proposed that this could well be seen as a primitive case of mental representation (Brüntrup & Jaskolla, 2016; Bohm, 1990; Bohm & Hiley, 1993; Pylkkänen 2006)

The critic of Russellian panpsychism might respond that although the theory is compatible in principle with present and future science, it is unscientific because it is unfalsifiable. By this, they might propose that no current or future scientific observation could establish the existence or non-existence of quiddities, and so Russellian panpsychism cannot be subject to experiment. Now, there are two ways our critic could make this claim. First, he could argue that any world in which Russellian panpsychism is true is *observationally equivalent* to one in which only entities traditionally depicted by the physical sciences exist. However, the issue with extending this consideration into a critique is that such negative reasoning is reversible; it is also the case that no current or future scientific observation could establish that all there is are structuralist material entities. In this manner, if the claim of observational equivalence holds, physicalism must suffer the same fate as Russellian panpsychism. Furthermore, it might not be the case that the two are observationally equivalent. At face value, they might differ on their observation of and ability to predict consciousness: Russellian panpsychism predicts that I should be conscious, whilst physicalism, in virtue of its Hard Problem, might not. In this way, the fact

that I am conscious is an observational distinction that might actually favour Russellian monism over physicalism from the offset.

The second way our critic might make the case that Russellian panpsychism is unfalsifiable is with the claim that there is no *observational data* that would directly support the claim that electrons have consciousness. I can certainly sympathise with this option: although there is also no data to support that electrons do not have consciousness, since Russellian panpsychism requests an amendment to the standard worldview, it should provide a hard datum in favour of its claims. However, Russellian panpsychism does have a hard datum that counts in favour of it, one in which we have always been acquainted with: our own conscious experience (Goff, 2017: 171-172). The reality of our consciousness is more evident to us than any empirical data, and so should be treated with the importance it deserves. Although the existence of conscious experience doesn't necessarily entail the truth of Russellian panpsychism, if the doctrine is internally stable, it becomes the best available theory within the Mind-Body debate, and we have evidence for the reality of its prediction and so the experience of electrons; the Russellian panpsychist could then extend the reality of consciousness as support for the reality of quiddities. In this manner, Russellian panpsychism has the *potential* to be immune to the charge that it apparently lacks observational data. In light of this reasoning, it appears that the best option for the critic of the doctrine is to attempt to evidence its internal incoherence instead.

2.3.2 The Combination Problem

The critic's best bet to argue for Russellian panpsychism's internal incoherency is the Combination Problem, or the problem of accounting for how quiddities combine to create macro-experience. This issue found its name through William Seager (1995) and is typically understood to have originated in William James's discussion of mental combination in *The*

Principles of Psychology (1890), and Lucretius's commentary on subjects in *On the Nature of Things* (1910). The Combination Problem dominates contemporary discussions about Russellian panpsychism and is accepted in its reality by nearly all of the doctrine's proponents and critics: Chalmers writes that it is the hardest problem faced by any Russellian-type view (Chalmers, 1996: 307), and Seager writes that it is the most difficult issue faced by any panpsychist theory (Seager, 1995: 280).

While sustained theoretical responses to the problem have been proposed (e.g., Roelofs, 2019; Miller, 2018), none are typically considered to have fully resolved it, and some commentators believe the problem is itself irresolvable (Whitehead, 1929a: 32; James, 1909: 213). Now, the Combination Problem must be resolved if Russellian panpsychism is to be theoretically sound and so forms its most pressing issue. However, considering all the virtues the doctrine has the potential to provide, if it can resolve this issue, it would become the best available account of consciousness; it would harness the virtues of both physicalism and dualism, avoid their vices, and be free of any major theoretical vice itself (Goff, 2017; Chalmers, 2016a; 2016b). In fact, if it is to solve *all* of its Combination Problems, Russellian panpsychism might simply be true.

In its most general form, the Combination Problem is the question of how micro-experiential quiddities combine to make macro-experience. This question separates into a variety of more complex forms (Chalmers, 2016b). We can delineate these into two broad categories: **Internal Combination Problems** and **Bridging Combination Problems** (Coleman, 2016). Internal Combination Problems are philosophical problems which focus upon whether quiddities are *theoretically capable*, viz. capable in principle, of combining to make macro-experience. Of these variants, the two most pressing, and of interest to this thesis are the Subject Summing Problem, and the Boundary Problem. Bridging Combination Problems, on the other hand, are

the scientific problems of what *conditions* would be needed for quiddities to combine to make macro-experience; they are the challenge of establishing the scientific compatibility of one's response to the Internal Combination Problems. In this manner, resolving the former category of problem logically precedes the latter: it is no good attempting to ground Russellian panpsychism in science if the theory is incapable of resolving its internal issues.

By 'combination,' I henceforth refer to the following:

Combination_{df}: some property instance or part P, is a combination of property instances or parts $p_1 \dots p_n$, iff:

- (1) P is fully grounded in, and exists in virtue of, and as nothing over and above $p_1 \dots p_n$, the real relations R that exist between them. And,
- (2) P is fully explained by $p_1 \dots p_n$, and those real relations R.³¹

We can say then that some property or part is a combination of some other properties or parts if those other properties or parts and the ways in which they are connected are sufficient to ground and explain away the property or part in question.

The first of our Internal Combination Problems is the **Subject Summing Problem**, or the issue of whether quiddities are theoretically capable of combining_{df} to make macro-experience. This line of worry stems from the following canonical statement from William James:

Where the elemental units are supposed to be feelings, the case is in no wise altered. Take a hundred [feelings], shuffle them and pack them as close together as you can (whatever that may mean); still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the

³¹ Although not directly relevant to the Combination Problem, for the sake of accuracy, I structure (2) in this manner so as to permit that weakly emergent and intelligibly emergent properties can satisfy this definition.

other feelings are and mean. There would be a hundred-and first-feeling there, if, when a group or series of such feelings were set up, a consciousness belonging to the group as such should emerge. And this 101st feeling would be a totally new fact; the 100 feelings might, by a curious physical law, be a signal for its creation, when they came together; but they would have no substantial identity with it, nor it with them, and one could never deduce the one from the others, nor (in any intelligible sense) say that they evolved it (James, 1890: 160).

James's worry is that no group of individual experiences seems to necessitate a further composite experience, since experiences are intuitively not of the right kind to form aggregates. Now, his reasoning here is grounded in a general resistance to the idea of *anything* combining (ibid: 158-159; Goff, 2016: 287-288). For present purposes, however, let us just consider the notion when applied to experience and so quiddities. The property of experience seems to require a 'subject' as a prerequisite. By this, I refer to the idea that any instance of experience appears to necessitate that it be realised by something, and that something must intuitively be an 'experiencer' in the sense that the experience is 'for it.' Since Russellian panpsychism proposes that micro-experience combines to make macro-experience, they must then also uphold that there are micro-subjects which can combine to make a macro-subject. Like macro-experience, a macro-subject is the unified form of subjectivity that we find within our own experience, and possess a pre-theoretical commitment to (Goff, 2017: 144).

The worry here, is that subjectivity is inherently individual and unique. The 'something it is like to be me' is not the 'something it is like to be' another, and the divide between myself and other distinct subjects appears impassable. In this sense, the problem of combining experiences – as required by Russellian panpsychism – necessitates a problem of combining subjects, and this notion appears to suffer the issue of contradicting our intuitive understanding of subjectivity. Arguably then, it is untenable for some particular number of micro-experiential

properties to add up to that with which we are acquainted, because individual experiences do not seem capable of being ‘summed;’ we might be able to ‘stack’ them, but we cannot ‘pool’ them (Goff, 2006: 53-59; Coleman, 2012: 147).

To illustrate this worry further, let us consider the following analogy by Daniel Stoljar (2006), as developed from a thought experiment by Ned Block (1980: 268-306). We are asked to imagine a race of tiny conscious aliens that are small enough and relevantly capable to realise and perform the causal powers and relations we understand to exist in our universe. In particular, we should imagine that they are capable of playing the role that quiddities are proposed to play, and so realise and replicate the structures and relations instantiated by micro-physical entities. For some unknown reason, these little aliens decide to do just that, and fly around in relevantly small spaceships, duplicating the exact casual description of fundamental entities. The idea here is that although the aliens are themselves conscious, it remains unclear how this consciousness could possibly ‘pool’ together into the consciousnesses of the higher-order entities realised by the relations and powers of which they instantiate. In this manner, it is allegedly unclear how quiddities can be expected to do the same (Stoljar, 2006: 120).

Now, while this analogy is useful, a few clarifications about it must be made. As noted, the Subject Summing Problem assumes that experience requires some form of subject. I shall grant this assumption. However, it must be noted that the theoretical subject as required in the case of quiddities should be starkly distinguished from the subjecthood of higher-order conscious entities, such as humans, or little conscious aliens. As emphasised, I adopt panexperientialism, not pancognitivism. In this manner, any subjecthood enjoyed by fundamental entities about micro-experiential properties does not entail any *thought* about that experience: there is an incredibly rudimentary ‘something it is like’ to be an electron, and so some relevantly

rudimentary notion of being experienced upon, but that something is not available to our electron in any introspective capacity; introspection falls into the higher-order category of thought, and so should be explained away at higher levels of abstraction. For the little alien, however, the ‘something it is like’ to be it would surely be accompanied by thought about that something, for, indeed, it decided to fly around in a little spaceship and play the role of fundamental properties. Now, combining thinking subjects is intuitively more difficult than combining non-thinking subjects, for thought about one’s subjectivity seems to sharpen the notion subjectivity about oneself, and thus delineates oneself from other subjects in a way that wasn’t available to the non-thinking subject. Thus, while ‘little aliens’ do help emphasise the issue with our combination relation, the theoretical issue of their combination should not be taken as accurately representative of the issue of combining quiddities.

Notwithstanding this reasoning, the problem still stands. Russellian panpsychism must account for whether quiddities can theoretically combine in the relevant sense so as to create composite experience. Indeed, there seems to be a sense in which this problem forms a similar epistemic gap to the Hard Problem for physicalism. Physicalism suffered the issue of explaining how non-experiential properties combine to create an experiential subject: how macro-experience combines out of non-experiential parts. The Russellian panpsychist now suffers the similar issue of explaining how micro-experiential parts constitute create macro-experience. Not only is this then a prudent theoretical issue that the Russellian panpsychist must resolve to establish the internal coherence of her doctrine, but its resolve might also be necessary to motivate her thesis over physicalism. Just as my reasoning for discounting emergentist Russellian panpsychism centred on the notion of providing some distinct theoretical advantage over physicalism, so too must our constitutive variety now assert itself positively lest the same charge against it stand: if quiddities do not transparently explain macro-experience, why not

just maintain that macro-experience comes from nonconscious microphysics in an equally puzzling way, and without the added metaphysical commitment of positing quiddities?

Our second Internal Combination Problem – the **Boundary Problem** – is most famously proposed by Gregg Rosenberg (2004) as an issue that one faces upon attempting to get around the Subject Summing Problem. This problem is intrinsically linked to the above reasoning concerning subjectivity. It is grounded in the fact that there appear to be very clear boundaries in conscious experience: some experiences are part of my own macro-experience, and most experiences – such as those had by other experiencing things – are not. Indeed, and as noted previously, it is this clear distinction between our own experience and everything else that delineates us as a subject (Rosenberg, 2004: 80). Typically, attempted responses to the Subject Summing Problem come in the form of a ‘combination principle,’ or a theoretical mechanism that accounts for *how* micro-experiences could combine to sum into a macro-experiential subject. However, in addition to this principle, one must also provide an account for how certain combined or macro-subjects remain distinct from others. The Boundary Problem for Russellian panpsychism is then the problem of accounting for how a combination mechanism that purports to resolve the Subject Summing Problem could also allow for the strict delineation of subjectivity we find in our own macro-experience.

The worry for the Russellian monist is that any mechanism which purports to allow for the combination of micro-experience must operate as a fundamental mechanism (since it governs fundamental parts). Without resolving the Boundary Problem, however, any fundamental combination mechanism about those micro-experiential parts might extend universally, such that our electrons, and every macro-consciousness all combine into one singular maximal consciousness. In this manner, by requiring a fundamental mechanism to account for the

combination of quiddities, the Russellian panpsychist might inadvertently render the entire cosmos as one large subject of experience whilst denying room for any delineation between oneself as a distinct subject, or potentially for oneself as a subject at all. The Russellian panpsychist must then arguably account for how micro-experience combines into macro-experience without necessitating that this combination relevantly extends universally towards the maximal level.³²

For the sake of clarity, we can state these problems as follows:

Internal Combination Problems:

- (i) *Subject Summing Problem:* Are quiddities theoretically capable of combining to form composite experiences?
- (ii) *Boundary Problem:* How can one resolve the former whilst sensibly constraining that combination mechanism's application?

If the Russellian panpsychist can resolve both of these problems, she may then begin to tackle the Bridging Combination Problems. Unlike Internal Combination Problems, there are no particularly developed accounts of this problem. Instead, they are usually portrayed broadly in terms of potential challenges to how Russellian panpsychism would be neurologically implemented or found compatible with other aspects of current science; they are the challenge of 'bridging' the theory with scientific theories (Coleman, 2016: 251). As noted, Internal Combination Problems are more pressing than Bridging Combination Problems since tackling the latter depends upon the resolving of the former: if one has not ironed out the theoretical issues in their theory, there is no point trying to implement that theory into contemporary

³² I say 'arguably' here, only to account for those Russellian panpsychists who might be willing to extend transitivity to this potentially unintuitive universal end (e.g. Roelofs, 2019).

science. Furthermore, since science is constantly developing and changing, changes in science could intuitively explain away the Bridging Combination Problems, so long as the theory has already proven its internal coherency. In this manner, internal problems have the theoretical precedence, and so logically precede the explanation of the other.

Following this relation of necessity, Bridging Combination Problems are less of a threat to Russellian monism: if one resolves the internal problems, and in light of no other theory of consciousness having established its coherence in this manner, it is likely that the bridging problems will be capable of being explained away. Due to this, and in light of the propensity of science to change, I shall not portray a strict negative bridging challenge for Russellian panpsychism, and instead opt to present it as a positive question:

Bridging Combination Problem: how, and under what conditions might quiddities combine to make larger forms of experience?

2.4 ~ Concluding Remarks

The remainder of this thesis will analyse whether Russellian panpsychism can offer a more theoretically promising solution to the Mind-Body Problem than physicalism. As we have seen, Russellian panpsychism avoids the dualist's vice of separating consciousness from the natural order and thus from problems of causation. However, she also agrees with the dualist that the properties described by current physical science are themselves insufficient to account for consciousness. By adding quiddities – fundamental properties which allegedly explain consciousness – she grounds consciousness at the fundamental level and so appears to avoid the physicalist's Hard Problem of consciousness, viz. of explaining *how it is* that things are conscious (Cutter, 2019: 113; Goff & Coleman, 2020: 304-305). While Russellian panpsychism appears to capture the virtues of both theories while avoiding their vices, it suffers its own

potentially damming problem – the Combination Problem. I have noted that unless this problem is resolved, as the debate stands currently, it is unclear whether we have sufficient motivation to adopt Russellian panpsychism over the currently dominant physicalist approach: if quiddities do not transparently explain macro-experience, why not just maintain that macro-experience comes from nonconscious microphysics in an equally puzzling way, and without the added metaphysical commitment of positing quiddities?

While this thesis will not provide an explicit resolution to the Combination Problem, I shall construct a new argument which will suggest that even with this problem looming, Russellian panpsychism can still provide the most theoretically promising solution to the Mind-Body Problem. With this in mind, in Chapter 3, I will consider one major argument that is typically understood to favour Russellian panpsychism and oppose physicalism in this vein – the Conceivability Argument. If this argument holds, then Russellian panpsychism would provide the best available account of consciousness; as we shall see, physicalism would be false, or Russellian monism would be true. However, in Chapter 4, I shall argue that although this argument remains popular, it is of no help to the Russellian panpsychist. In fact, as long as the Combination Problem remains, the Russellian panpsychist is vulnerable to an analogous, equally threatening argument. However, I shall propose that this is a fault of this form of argument's application, as opposed to a fault within Russellian panpsychism itself. After illustrating this fault, I will therefore advise that Conceivability Arguments should be relinquished from the Mind-Body debate.

With this in mind, Chapter 5 will propose a new and improved argument for Russellian panpsychism, concerning the evolution of consciousness. Within this, I will assess and deny the claim that the existence of the Combination Problem undermines the motivation for

Russellian panpsychism over physicalism: I shall demonstrate that even with the Combination Problem unresolved, Russellian panpsychism still provides the more attractive theory. In this manner, I shall provide a ‘safety net’ for any future responses to the Combination Problem, such that Russellian panpsychism can still be theoretically preferred over physicalism, even with its Combination Problem remaining unresolved.

3

The Conceivability Argument

Against Physicalism, and for Russellian Panpsychism

The chapter will introduce one argument that is popularly appealed to by Russellian panpsychists to assert their theory's theoretical advantage over physicalism – the Conceivability Argument. 'Conceivability Arguments' are commonplace within discussions concerning the Mind-Body Problem. The most prominent version of this argument endorses the claim that it is conceivable that phenomenal consciousness – 'Q' – is not entailed by physical properties – 'P' – and so it is metaphysically possible that P and not Q, and so physicalism is false (Chalmers, 1996; Jackson, 1998; White, 1986; Robinson, 1993; Nagel, 1974; Kripke, 1972; Goff, 2017). Of these versions, David Chalmers' 2-Dimensional Conceivability Argument has become perhaps the most prominently utilised within academic philosophy (Chalmers, 1996, 2010). In this chapter, I will expound Chalmers' simple Conceivability Argument, and then explain its more complex 2-Dimensional form in detail. I will then address some potential physicalist rebuttals to this argument, alongside the popular reasons for why Russellian panpsychists understand their doctrine to be invulnerable to such forms of argumentation. Following the conclusion of the Conceivability Argument, I will conclude that if the argument is sound, then either physicalism is false, or Russellian monism is true.

3.1 ~ The Conceivability Argument against Physicalism

One of the most popular ways in which the Russellian panpsychist argues for her theory's superiority over physicalism is via the Conceivability Argument. This form of argument harnesses the Hard Problem for physicalism into a conceivability claim and then links this conceivability to a claim about metaphysical possibility. This metaphysical possibility allegedly then provides an ontological conclusion which establishes physicalism's falsity, or Russellian monism's truth, solely via a priori means.³³ Although many believe that the causal closure arguments expounded against dualism in Chapter 1 support physicalism, since we do not yet know *all* of the empirical facts about consciousness, a posteriori methods are not yet available to provide conclusive evidence for any specific philosophy of mind hypothesis. In virtue of this, and because of the Conceivability Argument's current accessibility and strong conclusion, it is popularly appealed to and developed upon by critics of physicalism and proponents of the Russellian monism (e.g., Chalmers, 1996; 2010; Goff, 2017). In this subsection, I will introduce an initial 'simple' formulation of the Conceivability Argument against physicalism, before delving into Chalmers' more 'complex' and nuanced 2-Dimensional version in the next.

3.1.1 The Simple Conceivability Argument

In its contemporary guise, the simple conceivability argument against physicalism begins by appealing to the notion of *philosophical zombies* (henceforth, p-zombies). Unlike the zombies that we are acquainted with on television and in literature, p-zombies refer to an entity which is physically (and behaviourally) identical to ourselves, but that lacks consciousness entirely (Chalmers, 1996: 94; Goff, 2017: 78-79). At the global level, we can consider a p-zombie

³³ Other a priori arguments against physicalism have been proposed, such as by Jackson (1982), and Robinson (1982; 2016), to name but a few.

world: a world physically (and behaviourally) identical to ours, but where everyone is a p-zombie.

Let us entertain the notion of what this scenario might entail. Building from the fact that p-zombies are physically identical to ourselves, we can say that my p-zombie twin is molecule for molecule identical to myself in all physical capacities. We can also posit that that he is situated within an identical physical environment and social scenario (Chalmers, 1996: 94). Granting this, my p-zombie twin would be behaviourally and functionally identical to me: he would receive the same sort of information, process it in an appropriate and cognitively indistinguishable way to myself, and produce indistinguishable resultant behaviour. Like myself, he is tapping away on his computer, sipping coffee, and introducing philosophical notions. However, unlike myself, there is no phenomenal experience of pressure under my p-zombie twin's fingertips, no fragrant satisfaction from his sipping of the coffee, nor any understanding of the notions he is introducing to his thesis. My p-zombie twin has no phenomenal experience, for *there is nothing it is like to be him*.

There is little reason to believe that p-zombies exist in the actual world. Indeed, I know I am not a p-zombie by virtue of my own phenomenal experience, and so extend that – at least under this planet's laws of nature – entities of a relevantly similar physical constitution and behaviour to myself also enjoy this experience. However, many uphold that p-zombies are at least conceivable, such that we can coherently imagine them without any logical contradiction (Goff, 2017; Chalmers, 1996; 2010; Campbell, 1970; Kirk, 1974, Kripke, 1980). Now, what this notion of conceivability actually entails will be the notion of the next subsection. For the moment, however, it will suffice to say that the notion is closely tied to apriority: something is

conceivable just in case it is a rationally coherent hypothesis that cannot be ruled out a priori (Cutter, 2019, 110; Chalmers, 2002).

One might respond that p-zombies are simply not a rationally coherent notion. However, to do so, one must demonstrate a contradiction in their description. For example, to illustrate the incoherence of a flying pig, a mile-high unicycle, or a nonconscious physical isomorph, one must first give some idea of where the contradiction in such notions lie. Indeed, one can simply combine their concepts of pig and flight to conceive of a flying pig. Or they can imagine a normal unicycle and extrapolate the situation till it is a mile high. If the criticiser of these notions cannot point out where the concept of a ‘unicycle’ and of being a ‘mile high’ contradict each other upon combination, he cannot discern any obvious contradiction in the overall notion of a mile-high unicycle (Chalmers, 1996: 96). Similarly, if one cannot point out where the concept of ‘an entity physical identical to ourselves’ and ‘nonconscious’ contradict each other, then one can discern no obvious contradiction in the overall notion of a p-zombie. In this manner, the burden of proof arguably lies on those who deny the notion’s coherence.

Like the combination of our concepts of pig and flight, proponents of p-zombies argue that there is no obvious contradiction in the notion of a nonconscious but physically identical entity to ourselves: it is rationally coherent. To bolster this claim, we can consider an entity with nonstandard realisations of my own functional organisation, and which is intuitively nonconscious.³⁴ Consider, for example, Ned Block’s ‘Chinese Nation’ (1980). Here, we are asked to imagine the inhabitants of China organising themselves into a causal organisation isomorphic to my brain. Each person should be imagined to be simulating the function of a

³⁴ While such an entity is not *identical* to us – such that identity theorists would permit its coherence – we shall modify this obviously nonconscious entity’s physical constitution to show that there is intuitively no physical or functional change that will conceptually entail consciousness production.

single neuron, and to possess a radio link corresponding to synapses so as to communicate to other people simulating other neurons. To further reinforce the example, let us propose that the population resides in a giant robot body, equipped with the ability to receive input data and produce output responses via methods isomorphic to our own nervous and sensory system. It seems unlikely that our Chinese Nation would somehow form a ‘group mind’ where the individual parts give rise to conscious experience. In this manner, the notion that this system would lack conscious experience appears to be coherent (Block, 1980; Chalmers, 1996: 97).

We can now alter the thought experiment such that each person is replaced one by one with silicon replicas of neurons, each possessing the same ability to communicate with other silicon neurons. We should further imagine that all of these are then arranged in a silicon body which is entirely isomorphic in constitution and capacity to our human body. It is still unclear whether such a silicon isomorph would become conscious, as there is no clear point at which consciousness should emerge in contrast to our unconscious Chinese Nation. Now, whether this silicon isomorph would actually be or come to be conscious is unimportant to the analogy. What is important is that its lacking consciousness appears to remain a coherent possibility. In this manner, it appears possible to conceive of conscious experience not being entailed by these facts about physical, functional and behavioural organisation (ibid).

Given that it is conceptually coherent that my silicon isomorph is not conscious, proponents of p-zombies argue that a non-conscious physical isomorph is equally coherent. Indeed, one could replace each of the above silicon neurons one by one with real neurons, but this still does not seem to produce any conceptual entailment that the resultant entity must be conscious. If one accepts this reasoning, they must find the notion of p-zombies rationally coherent – at least

upon initial consideration. With this in mind, we can structure the first premise of our initial simple Conceivability Argument:

P.1 It is conceivable that there are p-zombies.

Now, as noted previously, p-zombies are probably not naturally possible in this universe, viz. given the known natural laws which govern our universe it is likely that anything physically identical to me is also conscious. However, since the notion of them appears rationally coherent, it seems possible that they *could have* existed: their conceivability appears to suggest their possibility. For example, it might be argued that an omnipotent creator could have created a p-zombie world if they so wished, or that upon creating some physical universe, had to ‘add in’ something else to ensure that that universe contained consciousness. In either case, it appears that their apparent conceivability entails a link to their possibility: although p-zombies are not possible in our world, they could occur in some other *possible* world, viz. their existence is a metaphysical possibility. With this in mind, we can structure the second premise of our initial simple Conceivability Argument:

P.2 If it is conceivable that there are p-zombies, it is metaphysically possible that there are p-zombies.

If it is metaphysically possible that our physical isomorph could exist without consciousness, one can then infer that consciousness must be in some way nonphysical. Now, as described in Chapter 1.3, grounding entails metaphysical necessitation:

Necessitation_{df}: $X_1 \dots X_n$ necessitates Y when the conditional, ‘if $X_1 \dots X_n$ then Y ’ is metaphysically necessary.

If facts $X_1 \dots X_n$ ground fact Y , then it is metaphysically necessary that if $X_1 \dots X_n$, then Y . By this logic, any grounded fact or entity is necessitated by the obtaining of its ground (Cutter, 2018: 45). Accordingly, if physical facts ground and so necessitate the reality of consciousness, it would be impossible that there could exist a world which is physically identical to ours but which there is no consciousness (Goff, 2017: 79). Since the metaphysical possibility of p-zombies shows that such a world could exist, then it cannot be the case that the physical facts ground and so necessitate the reality of consciousness: consciousness must be something over-and-above the physical components of the universe. Accordingly, the metaphysical possibility of p-zombies would entail the non-physicality of consciousness. In this manner, we can structure the third premise of our initial simple Conceivability Argument:

- P.3 If it is metaphysically possible that there are p-zombies, then consciousness is nonphysical.

With these premises in place, we can produce an initial formulation of the simple Conceivability Argument:

- | | |
|-------|--|
| P.1 | It is conceivable that there are p-zombies. |
| P.2 | If it is conceivable that there are p-zombies, it is metaphysically possible that there are p-zombies. |
| P.3 | If it is metaphysically possible that there are p-zombies, then consciousness is nonphysical. |
| <hr/> | |
| C. | Consciousness is nonphysical. |

With the basic argument set out, we can now formalise it into a more precise simple Conceivability Argument to be directed against physicalism. ‘P’ should be taken henceforth to

represent the complete pure physical truths about the universe. This includes truths about all of the physical entities, properties, relations, and laws (including those known, and those currently not known but which nonetheless exist throughout space and time) that can be captured via the mathematico-nomic vocabulary of physics (Goff, 2017: 79-80). ‘Q,’ should be taken henceforth to represent all phenomenal truths about that universe. ‘&’ should be taken henceforth to represent ‘and,’ and ‘~’ should be henceforth taken to represent the negation operator ‘not.’ We can formulate our precise simple Conceivability Argument as follows:

P.1	It is conceivable that $P \& \sim Q$.
P.2	If it is conceivable that $P \& \sim Q$, it is metaphysically possible that $P \& \sim Q$.
P.3	If it is metaphysically possible that $P \& \sim Q$, then physicalism is false.
C.	Physicalism is false.

(Chalmers, 2010: 108).

In anticipation of our more complex Conceivability Argument, it will be helpful here to reinforce these premises in a little more detail. Now, P.1 proposes that there exists an epistemic gap between the physical and the phenomenal and denies that any epistemic relation between them could be known or explained a priori (Chalmers, 2010: 109). In this manner, P is proposed to not epistemically entail Q, such that Q is not deducible from or fully explainable by P. This premise harnesses the physicalist’s Hard Problem into a claim about conceivability. If it were the case that P epistemically entailed Q, it would be the case a priori that if P then Q; complete knowledge of P would therefore also provide complete knowledge of Q. Since one can arguably conceive of P without Q, then it appears the case that ‘if P’ does not epistemically entail ‘then Q’. In this manner, P.1 establishes an epistemic gap between the domains.

After establishing a lack of epistemic entailment between P and Q, P.2 then proposes a failure of ontological entailment, viz. the possibility of an ontological separation between the natures of P and Q. We can say then that where P.1 establishes an epistemic gap, P.2 establishes an ontological one. Now, if it were the case that P ontologically entailed Q, it must then necessitate Q such that ‘if P then Q’ holds universally; ‘P&~Q’ would be an incoherent notion. However, since one can conceive of P without Q – and we have not established any incoherence in the conceived scenario – then it seems that it could have been the case that P and not Q, and so P must not (entirely) necessitate Q. In this manner, the fact that there is an epistemic gap between P and Q suggests that there is an ontological gap between them. Accordingly, the argument infers from the epistemic conclusion of conceiving of P&~Q to the metaphysical possibility of P&~Q.

As expounded in Section 1.3, physicalism requires that P necessitates *all* truths. More specifically, P is proposed to constitutively ground everything, including Q. What this means, is that if P constitutively grounds Q, then if P obtains, Q *necessarily* also obtains: P necessitates and will always necessitate Q, in any possible world (Goff, 2017: 79; Rosen, 2010, 109-63; Fine, 2012; Trogon, 2013, 465-85). If this was to hold, then *any* minimal physical duplicate world to ours must necessarily have consciousness. However, our p-zombie world is a world that is physically identical to ours, but which does not have consciousness. Where P.1 showed that this world is conceivable, P.2 has shown that such a world is metaphysically possible: it is possible that there could be phenomenal truths Q that P does not necessitate (Chalmers, 2010: 110). This possibility is inconsistent with physicalism’s account of grounding, and so by *reductio*, one reaches the conclusion of physicalism’s falsity.

Now, the physicalist may object that this Conceivability Argument in a number of ways. On the one hand, our type-c physicalist might contend that we simply cannot imagine the physical complexity required to decipher any claim about the necessitation between P and Q. For example, there are an estimated hundred billion neurons in the human brain (Herculano-Houzel: 2009). Since most of these remain undocumented, we are arguably in no position to sit and imagine them in their entirety such that we can decipher claims of entailment about them. In this manner, he might claim that P.1 is flawed, and so P.2 and P.3 cannot be established so as to genuinely threaten the claim that ‘if P then Q.’ On the other hand, our type-b physicalist might argue that the adopted notion of conceivability is an imperfect guide to possibility, or indeed that conceivability as a whole is an imperfect guide. Here, the physicalist might claim that since the identity between P and Q is epistemically primitive, it will always be conceivable that $P \& \sim Q$. However, although there is an epistemic gap between P and Q, since it is not an ontological gap, it is still not ontologically possible that $P \& \sim Q$ regardless of this conceivability. These criticisms do indeed raise valid concerns for our simple Conceivability Argument. However, as noted, the above argument is a *simple* one, expounded only to introduce the basic argument structure and notions relative to its more complex formulation. In the following subsection, I will set out a complex form of the Conceivability Argument, which introduces a more precise notion of conceivability, and which allegedly bypasses these concerns so as to genuinely threaten the physicalist hypothesis.

3.1.2 The Complex Conceivability Argument

The complex Conceivability Argument adopted in this thesis utilises the criteria set out by Chalmers’ ‘2-Dimensional Conceivability Argument’ in *The Conscious Mind* (1996), and *The Character of Consciousness* (2010). Here, Chalmers develops a nuanced and specialised account of conceivability to input into our simple Conceivability Argument. This account of

conceivability is intended to strengthen P.1 and P.2 of our argument, so as to produce a more robust argument against physicalism, whilst also bypassing allegations about the inconceivability of $P \& \sim Q$, and of conceivability as an imperfect guide to possibility. In this subsection, I will introduce each of the qualifications within this specialised account, and then input them in turn into P.1. Once all of the qualifications are set out in this way, I will input them into a Complex Conceivability Argument and explain the threat this finalised argument delivers to physicalism, alongside how it supports Russellian monism.

The simple Conceivability Argument suggested that conceivability should be understood in epistemological terms, whereby some hypothesis X was conceivable just in case it is a rationally coherent hypothesis that cannot be ruled out a priori. Harnessing this initial formulation, Chalmers adds three additional qualifications. Now, as considered in the previous subsection, human reason is contingent and ‘knowledge’ is a work in progress, and so it is highly unlikely that any person’s current cognitive makeup and knowledge of the world could allow them to truly conceive of P (where P continues to represent the complete pure physical truths about the universe). For this reason, Chalmers introduces his first qualification: the distinction between **prima facie** and **ideal** conceivability.

We can say that X is *prima facie* conceivable when the subject is unable to rule out the hypothesis expressed by X by a priori reasoning upon initial consideration (Chalmers, 2010: 144). Examples of this form of conceivability might include statements such as “it is conceivable that ‘ $892 \times 98 = 87216$,’ ‘ $783 \times 1234 \neq 966222$,’ or ‘there are cousins but no siblings.’” For most people, each statement presents a *prima facie* conceivable scenario: the first example appears to present an answer that is in the right ballpark, and most cannot initially rule it out as false due to their own cognitive limitations; the second example also presents an

answer that might be *prima facie* in the right ballpark, as most cannot rule out that 966222 is *not* the correct answer upon initial consideration; the final example might *prima facie* present a coherent state of affairs, for initial consideration might tell us that siblings don't necessitate cousins. On the other hand, statements such as ' $1 + 3 = 5$,' 'there is a square circle' or 'his mother is a married bachelor' are not *prima facie* conceivable, as one can immediately rule out their truth by *a priori* reasoning.

We can say that X is ideally conceivable when the subject is unable to rule out the hypothesis expressed by X by *a priori* reasoning even upon ideal rational reflection (Chalmers, 2010: 144). Ideal rational reflection refers to a perspective of epitomal reasoning, whereby some statement is only ideally conceivable if *no amount* of *a priori* reasoning could deduce its falsity. Where *prima facie* conceivability was linked to a subject's cognitive limitations, ideal conceivability abstracts away from those limits. For example, where ' $892 \times 98 = 87216$ ' was *prima facie* conceivable for most, it is certainly not ideally conceivable. Ideal rational reflection would allow me to rule out the statement, as 892×98 is 87416, and this is an *a priori* fact about the world: it is not possible for it to be otherwise. Similarly, while the statement ' $783 \times 1234 \neq 966222$ ' might be *prima facie* conceivable, ideal reflection tells me that 783×1234 *is* 966222, and so its negation is not ideally conceivable. Finally, while 'there are cousins, but no siblings' appeared *prima facie* conceivable, it is not ideally conceivable: upon reflection, to be a cousin demands that you have parents which are siblings, and so if there are cousins, there must be siblings (Goff, 2017: 81). In contrast, statements such as ' $1 + 3 = 5$,' 'there is a square circle' or 'his mother is a married bachelor,' are neither *prima facie* conceivable, *nor* ideally conceivable.

Of these two notions, it is ideal conceivability that is of interest to our complex Conceivability Argument. The reason for this is that ideal conceivability is a better guide to possibility than *prima facie* conceivability. For example, in the above examples, we saw that ‘ $783 \times 1234 \neq 966222$ ’ is *prima facie* conceivable: we can *prima facie* conceive of the negation of a mathematical truth. However, since the statement ‘ $783 \times 1234 = 966222$ ’ is necessarily true, it is not possible for its negation to obtain. In this manner, there is good reason to think that *prima facie* conceivability is not a good guide to possibility. As noted, *prima facie* conceivability is dependent upon one’s cognitive capacity and limitations. Ideal conceivability, however, abstracts away from such limitations in such a manner that it appears to track more precisely onto possibility: it is not ideally conceivable that ‘ $783 \times 1234 \neq 966222$,’ since ideal rational reflection tells me that 783×1234 is and can only be 966222, and so it cannot be the case that ‘ $783 \times 1234 \neq 966222$ ’ in any possible world.

With this qualification made, we can amend P.1 such that the statement ‘It is conceivable that $P \& \sim Q$,’ now means:

‘It is ideally conceivable that $P \& \sim Q$.’

For it to be the case that $P \& \sim Q$ is ideally conceivable, it must be the case that even from a perspective of epitomal reasoning, no amount of a priori reasoning could deduce that $P \& \sim Q$ is false. What this requires, is that even with an idealised knowledge of the variables P and Q , the epistemic gap between them will not be explained away, and so that the existence of P without Q will remain coherent. The proponent of Conceivability Arguments proposes that it is indeed ideally conceivable that $P \& \sim Q$. As we have seen with regards to the Hard Problem, it is intuitively not the case that P a priori entails Q , as an epistemic gap is firmly established between them. As already argued, the vocabulary of P appears incapable of capturing the

distinctive nature of Q, as everything we learn from explanations regarding P are consistent with an absence of Q. For this reason, while it is clearly *prima facie* conceivable that $P \& \sim Q$, one can allegedly extrapolate this notion up to some ideal notion of conceivability: no matter the progression of physicalism, by virtue of the Hard Problem it will never be the case that Q is explained away by P for P is simply not equipped to explain away Q. Since ideal conceivability of this kind is a good guide to possibility – in that it abstracts away from human cognitive limitations – we can strengthen the claim of P.2: that it is metaphysically possible that $P \& \sim Q$.

The above examples of *prima facie* and ideal conceivability link closely to Chalmers next qualification to the notion of conceivability: the distinction between **positive** and **negative** conceivability. Now, each of the examples set out above are all versions of negative conceivability: they are defined in terms of what a subject can *rule out* a priori (Chalmers, 2010: 144). In this manner, some statement X is *prima facie* negatively conceivable when our subject cannot rule out X by a priori reasoning upon initial consideration. Similarly, some statement X is ideally negatively conceivable when our subject cannot rule out X by a priori reasoning even upon ideal rational reflection.

Positive conceivability, on the other hand, is defined in terms of what a subject can *form a positive conception of*: some statement X is positively conceivable when one can coherently imagine a situation in which X obtains. Some statement X is *prima facie* positively conceivable when our subject can imagine a situation that seems to be initially coherent and within which, X appears to obtain. For example, I can *prima facie* positively conceive of the statement ‘there is a flying pig,’ but probably not the statement, ‘there is a 1-billion-sided shape.’ In the first instance, I can *prima facie* positively imagine a situation whereby I combine my concepts of

‘flying’ and ‘pig’ to positively imagine a pig which is flying. I probably cannot, however, *prima facie* positively imagine a 1-billion-sided shape, as no positive conception of this notion is available to my imagination upon initial reflection. Conversely, some statement X is ideally positively conceivable when ideal rational reflection reveals the imagined situation as one in which S is the case (*ibid*). In this manner, the statements ‘there is a flying pig,’ and ‘there is a 1-billion-sided shape’ might both be ideally positively conceivable: upon ideal rational reflection the imagined scenario of a flying pig still obtains, and I *might* be able to fill in the arbitrary details of our shape (having 1 billion sides) so as to produce a positive conception, and without any contradiction revealing itself.³⁵

Chalmers proposes that either ideal positive *or* ideal negative conceivability can be used in our complex Conceivability Argument since both are good approximators to possibility (*ibid*: 147). Ideal negative conceivability has already been shown to be a better approximator to possibility than *prima facie* negative conceivability, and a similar case can be made for ideal positive conceivability: a 1-billion-sided shape was not *prima facie* positively conceivable due to our own cognitive limitations, but ideal positive conceivability might – in principle – abstract away from these limitations to show the overall coherence, and so possibility of the statement. In this manner, where *prima facie* positive conceivability can lead us astray when concerning the possibility of a statement, ideal positive conceivability might provide a better link to genuine possibility.

For the purposes of this thesis, I shall focus only on ideal negative conceivability. My reasoning is twofold. First, positive conceivability is a difficult notion to define, as even the ideal notion

³⁵ I stress ‘might’ here due to the ideally positive conceivability of this scenario being likely dependent upon our own cognitive limitations, as opposed to the coherence of the notion. While we might know intuitively that the scenario is coherent (via negative reasoning), it is perhaps unclear whether we could ever have a positive conception of it obtaining. I will return to this potential issue in the coming paragraphs.

of it remains dependent upon the nature of human imagination and their cognitive capacities: even from an ideal perspective, we will still only possess the same processing power we do right now, regardless of whatever arbitrary details have been revealed to our ideal perspective. Indeed, in the above example, it is still arguably unclear whether an idealised agent could positively conceive of a shape with 1-billion sides: although a billion is a concept we can understand, we might not necessarily be able to form it into a positive conception due to our cognitive limitations. This is not to say that its being inconceivable entails its impossibility. Indeed, since Chalmers gives us little reason to think our cognitive capacities should have any bearing on the limits of possibility, it is unclear how accurate of a possibility claim positive conceivability can deliver. Ideal negative conceivability, on the other hand, harnesses the notion of ideal rational coherence, in such a way that our restricted human cognitive capacities have less potential to misguide our possibility claims. Concerning our 1-billion-sided shape, I can harness my understanding of a billion, and of a shape, so as to deduce that there could be no contradiction in such a shape existing: we need not positively imagine some scenario X, but rather must simply know that it can't be ruled out.

My second reason for adopting ideal negative conceivability is due to the fact that we will be inputting our notion of conceivability into claims about P and Q. And, in the case of conceivability claims about $P \& \sim Q$, one might run into issues if they adopt ideal positive conceivability. To positively conceive of $P \& \sim Q$, one must be able to positively imagine a scenario in which P obtains but Q does not. Since I am a conscious thinking thing, any positive imagination I can conjure (whether *prima facie* or ideal) might then also entail Q by virtue of my possessing it.

One might respond that we can still positively conceive of an *absence* of Q in some counterfactual world regardless of our possession of Q: our own possession of Q need not contaminate our conception, for we do need not conceive of p-zombies from that entities' first-person perspective (or lack thereof). Whether or not this is the case is contentious, but unimportant to my reasoning. Since Chalmers' proposes that both ideal positive and ideal negative conceivability are good approximators to possibility, I believe that we will do well to adopt the least contentious variant if we are to establish a robust case against physicalism. Ideal negative conceivability about $P \& \sim Q$ is clearly an easier task since it only asks us to rule out the conceivability of $P \& \sim Q$ from a position of ideal reasoning, as opposed to asking us to positively (and ideally) imagine a scenario in which $P \& \sim Q$ actually obtains (Marcus, 2004: 477-90). In this manner, adopting ideal negative conceivability will provide a more accessible conceivability claim, with an equally strong link to possibility.

With this reasoning in mind, we can now further amend P.1 such that the statement 'It is conceivable that $P \& \sim Q$,' now means:

'It is ideally negatively conceivable that $P \& \sim Q$.'

As noted, the account of ideal conceivability expounded above was already a form of negative conceivability, and so the conditions for $P \& \sim Q$ being ideally negatively conceivable remain the same as those necessary for its being ideally conceivable: it must be the case that even from a perspective of epitomal reasoning, no amount of a priori reasoning could deduce that $P \& \sim Q$ is incoherent. The epistemic gap between them must remain impassable even with an idealised knowledge of P and Q. As considered above, there is a case to be made that we do indeed have this form of conceivability about $P \& \sim Q$: the vocabulary of P is intuitively of the *wrong* kind to ever explain away the epistemic gap as illustrated by the Hard Problem; as argued in Chapter

1, no matter how complex one's account of P with regards to Q becomes, the explanation might remain solely one of correlation, as opposed to one of epistemic entailment. Granting this, the possibility of $P \& \sim Q$ will never be able to be ruled out, even upon an idealised perspective about P and Q. Since (ideal) negative conceivability is a good guide to possibility, by amending P.1 in this way, we have again further strengthened the claim of P.2 that it is metaphysically possible that $P \& \sim Q$.

Chalmers' final qualification to our notion of conceivability centres on a problem introduced by Saul Kripke in *Naming and Necessity* (1980), and Hilary Putnam in *Mind, Language, and Reality* (1975) and *Meaning and Reference* (1973) (Chalmers, 1996: 56). Kripke and Putnam propose that there can be sentences which are necessarily true despite not being knowable a priori, and that these are cases of 'a posteriori necessities.' The common example appealed to when illustrating a posteriori necessities is the true statement (1): 'Water is H_2O .'³⁶ Now, one can neither deduce nor rule out whether water is a compound made up of two hydrogen atoms and one oxygen atom by a priori means: coming to know this required a posteriori methods. Nonetheless, these methods appear to have established that it is a necessary truth that water *just is* H_2O and is so in any possible world.³⁷

Although statement 1 appears necessarily true, since its truth is not knowable a priori, there seems to be a sense in which one might be able to conceive of its negation: 'Water is not H_2O ,' or ~ 1 . We cannot establish any contradiction in the negation of 1 via a priori means, and so ~ 1 appears conceivable via our definition of conceivability. However, while it might appear

³⁶ It is worth noting that although 'Water is H_2O ' is true, the converse statement ' H_2O is water,' need not hold universally: H_2O can also be water vapour, or ice.

³⁷ It could turn out that water is something other than H_2O , such as the more complex combination of micro-level parts that constitute that compound. However, for the sake of simplicity, let us use H_2O as a placeholder for whatever the essential makeup of water is, or may turn out to be.

conceivable, it is still arguably impossible for ~ 1 to obtain: it is an empirical truth that water *just is* H₂O and, therefore, anything ontologically identical to water must also be H₂O (Chalmers, 1996: 57). This identity relation must then also extend into any possible world, and thus it is not metaphysically possible that ~ 1 . If this is the case, then the a priori methods of conceivability discussed already appear to have no bearing on our possibility claim: ~ 1 is conceivable but *not* possible, and so it must be the case that conceivability – as currently defined – does not necessarily entail metaphysical possibility. Accordingly, P& $\sim$ Q's conceivability doesn't necessarily necessitate its metaphysical possibility either.

To illustrate this, let us consider some counterfactual world X. This world is exactly like ours in terms of P, but the colourless, transparent, odourless wet stuff that forms seas, lakes, rivers, and rain has a different chemical composition (Goff, 2017: 83). Let us call this composition X₅Y.³⁸ X₅Y has a different chemical nature and inconsistent chemical definition to H₂O but is superficially identical in terms of its macro-level characteristics. Arguably, even if we cannot positively conceive of this world, we cannot rule out that such a world could exist, and so we can establish a negative conceivability claim. Granting this, we have now arguably conceived of ~ 1 : we have conceived of a world in which water is something other than H₂O. Following our link between conceivability and possibility, one might therefore claim that it is metaphysically possible that ~ 1 .

However, since it is an a posteriori necessity that water just is H₂O, it cannot be metaphysically possible that ~ 1 : 1 is necessarily true, even in world X. Now, if one accepts a posteriori necessities of this form, then P.2 of our Conceivability Argument appears to be flawed:

³⁸ 'X' and 'Y' should be taken to stand as a placeholder for individual compositional parts that already exist within our account of P, and which can (unbeknownst to us) form a compound superficially identical to water.

regardless of our qualifications to conceivability, it is not necessarily the case that if a sentence is conceivably true, then it is possibly true, for it was conceivable that water was not H₂O, yet it was not actually possible. In this manner, the physical grounding of Q in P could also be an a posteriori necessity: although conceivable, P&~Q is not metaphysically possible. Accordingly, if one wishes to utilise Conceivability Arguments against physicalism, they must either deny the existence of a posteriori necessities or provide some further qualification to their notion of conceivability that accounts for this problem.

Following Chalmers, I shall grant that a posteriori necessities exist, and so introduce the final qualification to our notion of conceivability: that of **primary** and **secondary** conceivability, and their respective links to primary and secondary possibility (Chalmers, 2010: 144-150). This distinction is intended to illuminate why it is that ~1 appeared conceivable but was not possible, and will then be built into a defence regarding the conceivability and so metaphysical possibility of P&~Q.

The distinction between primary and secondary conceivability begins by distinguishing of two aspects of meaning within Kripke's *Naming and Necessity* (1980): primary and secondary intensions. According to Chalmers, the primary intension of a referent within a sentence corresponds to what Kripke refers to as the referent's *appearance*, viz. those appearance properties we use to 'pick the referent out' (ibid; Goff, 2017: 86). For example, the primary intension of water is its being the colourless, transparent, odourless wet stuff that forms seas, lakes, rivers, and rain. On the other hand, the secondary intension of a referent within a sentence corresponds to what Kripke refers to as the referent's *essence*, viz. those essential properties that make up what the thing is in-of-itself (Chalmers, 1996: 63). For example, the secondary

intension of water is that it is a compound made up of two hydrogen atoms and one oxygen atom.

Concerning our conceived world X, the primary intensions of 'Water is X_5Y ,' or 'Water is not H_2O ,' are approximated by the sentences, 'The colourless... wet stuff that forms seas... is X_5Y ,' and 'The colourless... wet stuff that forms seas... is not H_2O ,' respectively. On the other hand, the secondary intensions of 'Water is X_5Y ,' or 'Water is not H_2O ,' are approximated by the sentences, ' H_2O is X_5Y ,' and ' H_2O is not H_2O ,' respectively (Goff, 2017: 87). In this manner, ~ 1 is true in world X when we are thinking of the sentence in terms of its primary intension: the colourless... wet stuff is indeed X_5Y . However, ~ 1 is false in world X when we are thinking of the sentence in terms of its secondary intension: H_2O is not X_5Y as it is a different chemical compound with an inconsistent chemical definition. We can say then that in world X, X_5Y is just a compound which superficially resembles water in such a manner that it satisfies its primary intension, but not its secondary intension.

We can say that some sentence is primarily conceivable when one can conceive of a possible world in which its primary intension is true. Similarly, we can say that some sentence is secondarily conceivable when we can conceive of a possible world in which its secondary intension is true. Now, since the secondary intension of 'Water is X_5Y ' is captured by the sentence ' H_2O is X_5Y ,' we can see that 'Water is X_5Y ' is clearly false in all possible worlds. Water may be superficially the colourless, transparent, odourless wet stuff that fills seas and lakes here, but that stuff just is essentially the chemical compound H_2O , and so water just is H_2O and not X_5Y .³⁹ Following this reasoning, any conceivable situation in which water appears to be something other H_2O – such as in our counterfactual world X – should in fact be described

³⁹ See Evans (1979: 161-89) and Davies & Humberstone (1980: 1-30) for more details.

as a situation in which water is still H₂O, but that there also exists a watery-like-stuff which is not H₂O. Or, more precisely, any conceivable situation in which water appears to be something other than H₂O should be defined as a situation in which the secondary intension of ‘Water is not H₂O’ is false, but the primary intension of ‘Water is not H₂O’ is true. In this manner, we can say that ‘Water is not H₂O’ is primarily conceivable, but that it is not secondarily conceivable.

As we have seen, if some sentence is primarily conceivable, then there is a possible world in which its primary intension is true. Similarly, if some sentence X is secondarily conceivable, then there is a possible world in which its secondary intension is true (Chalmers, 1996: 64). In this manner, we can formulate the following two conceivability principles with regards to possibility:

Primary Conceivability Principle – If some sentence X is primarily (ideally negatively) conceivable, it is primarily possible that X.

Secondary Conceivability Principle – if some sentence X is secondarily (ideally negatively) conceivable, it is secondarily possible that X.

While both forms of conceivability entail forms of possibility, in the majority of cases, the primary and secondary intensions of a referent differ, and so the conceivability and possibility claims one can make about them differ accordingly. For example, in the case of ~1, the primary intension of water differs from the secondary intension, and so while one can intuitively achieve primary conceivability about it (and so primary possibility), one cannot achieve secondary conceivability (and so secondary possibility) about it.

Primary and secondary conceivability typically establish distinct forms of possibility. However, it is only secondary conceivability which translates to the relevant form of metaphysical possibility here (it is only secondary conceivability which corresponds to the relevant metaphysical possibility of the sentence) (Chalmers, 2010: 147; *ibid*, 1996: 63). The primary conceivability of ‘Water is not H₂O’ entails that ‘it is possible that there is a world in which that the colourless... wet stuff that forms seas... is not H₂O.’ This amounts to the possibility claim that there could exist something superficially like water, but which is not actually water. This does not entail any metaphysical claim about the actual nature of water, but rather simply about the appearance of our referent. On the other hand, the secondary conceivability of ‘Water is not H₂O’ would entail that ‘it is possible that there is a world in which ‘H₂O is not H₂O.’ While secondary conceivability of this kind was unattainable with regards to statement 1, if attained, it would establish a metaphysical possibility claim about the *essence* of water.

The distinction between primary conceivability and secondary conceivability should illuminate exactly why statement 1 appeared conceivable, but not metaphysically possible. Through utilising primary conceivability, one (typically) only achieves claims about primary possibility (Primary Conceivability Principle). Accordingly, sentences can be primarily conceivable (and so primarily possible), but not secondarily conceivable (and so not secondarily possible). In this manner, primary conceivability does not typically appear to entail secondary possibility, and so does not typically entail the relevant form of metaphysical possibility. Translated to statement 1, we can see that it was primarily conceivable that ‘Water is not H₂O’ and so primarily possible, but it was not secondarily conceivable that ‘Water is not H₂O’ and so it was not metaphysically possible.

While this distinction should illuminate *why* conceivability did not entail relevant metaphysical possibility in the case of our a posteriori necessity, Chalmers must now defend the claim that the (ideal negative) conceivability of $P \& \sim Q$ *does* entail the metaphysical possibility of $P \& \sim Q$. Now, as noted prior, whether or not a sentence is secondarily conceivable and so secondarily possible is reliant on empirical factors. In the case of water, these empirical factors established an a posteriori necessity, and so denied the secondary conceivability of statement 1. Unlike secondary conceivability, primary conceivability appears to fall into the a priori category, as it depends on matters of a priori reasoning (Chalmers, 2010: 146). Since Conceivability Arguments intend to solely use a priori reasoning to deduce physicalism's falsity, the relevant notion of conceivability within them refers to primary conceivability, not secondary conceivability. However, primary conceivability has been shown to typically only entail primary possibility. Since $P \& \sim Q$ is only relevantly possible if its secondary intension is possibly true, the primary conceivability of $P \& \sim Q$ might then fail to present any genuine metaphysical possibility of $P \& \sim Q$, and so fail to threaten physicalism.

To get around this issue, Chalmers refers to a further aspect of Kripke's work. As noted already, the primary intension used to distinguish a referent (the appearance properties) are typically distinct from the secondary intensions of a referent (the essential properties) (Chalmers, 1996: 146). While Kripke acknowledges that this is true of most scientific kinds, he proposes that this is not the case when it comes to consciousness. He writes, "Pain... is not picked out by one of its accidental [appearance] properties; rather, it is picked out by the property of being pain itself, by its immediate phenomenological quality. Thus pain, unlike heat, is not only rigidly designated by "pain" but the reference of the designator is determined by an essential property of the referent" (Kripke, 1980: 15). What Kripke proposes here, is that we think about pain in terms of *how it feels*, viz. how it is presented to us. However, while how it feels is an appearance

property, this property is not a non-essential aspect; pain is *just a feeling*. In this manner, how pain feels is both the appearance property and essential property of pain. Per Kripke, conscious states are defined solely in terms of what it is like to be them – unlike scientific kinds such as heat or water, the appearance properties and the essential properties of Q arguably cannot be distinguished from one another.

We can summarise Kripke's reasoning as the thesis that terms expressing conscious states lack distinct primary and secondary intensions. Concerning $P \& \sim Q$, we can say then that sentences involving terms about Q lack distinct primary and secondary intensions (unless, of course, those sentences also involve other terms that have distinct primary and secondary intensions) (Goff, 2017: 89). If the sentence $P \& \sim Q$ lacks distinct primary and secondary intensions in this way, and we have (ideally negatively) primarily conceived of $P \& \sim Q$, we can therefore infer from its primary intension that its secondary intension is possibly true, and so that it is secondarily and so metaphysically possible. In this manner, we can bypass the typical empirical requirements for secondary conceivability, and reach it through a priori consideration about the primary intensions of Q.

Now, for this to obtain, the *entire* sentence $P \& \sim Q$ must lack distinct primary and secondary intensions. And, although the intensions of Q might coincide, the case is much less clear for P. In the case of P, let us consider 'mass.' Intuitively, the primary intension of 'mass' is tied to a certain theoretical 'mass' role (such as resisting acceleration and exerting attractive forces on other things with mass). However, the secondary intension of 'mass' might be tied to whatever the property is that actually plays the role. In this case, if property M plays the mass role in this world, then one might hold that in any other world w in which mass is instantiated, mass is M . If there are some other worlds in which some other property M' plays this role, then M' is not

mass in w (it is, at best, ‘pseudomass’).⁴⁰ Now, if property M plays the mass role in this world, then mass is only actually instantiated in worlds with property M playing the mass role (Chalmers, 2010: 150). In this manner, there could be a possible world with properties resisting acceleration and exerting attractive forces in a superficially similar way to mass, but since they do not have property M playing those roles, it is not really a world with mass.

If it is the case that P possesses distinct primary and secondary intensions in this way, then one cannot as easily infer the link from primary possibility to secondary possibility in $P \& \sim Q$. World X could have the same structural profile as physics in the actual world (in terms of its primary properties), but also a different intrinsic or essential profile (in terms of its secondary properties). In this manner, it could be that $P \& \sim Q$ is primarily possible, but still not secondarily possible, as ‘ P ’ would entail terms which have distinct primary and secondary intensions. Granting this, the truth of P.2 of our simple Conceivability Argument would not be established.

Chalmers notes that for this line of reasoning to hold, P and Q must be related in a certain specific way. Specifically, this requires that while it is primarily possible in some world that $P \& \sim Q$, it is not secondarily possible in *any* world that $P \& \sim Q$. For this to obtain, it must first be the case that a world may verify P without satisfying P . Per Chalmers, to verify P , a world must have the right structural profile, while to satisfy P , a world must have the right structural and intrinsic profiles (Chalmers, 2010: 151). If w is a world in which the role of mass is played by properties other than mass, which are otherwise distributed in the right way over spacetime and have appropriate causal and nomic relations among them, then w will verify P but will not

⁴⁰ It is worth noting that for people who think properties are powers, (e.g. Mumford (2004), there is no distinction between the property and the role. In this case, P ’s primary and secondary intensions would coincide, such that secondary conceivability can be maintained about $P \& \sim Q$.

satisfy P. Here, the physics of w has the same structural profile to our physics, but has a different intrinsic profile, in that it differs in the intrinsic properties that fill this structure.

If a world can verify P without satisfying P in this way, it may be primarily possible that $P \& \sim Q$, but not secondarily possible, and so that P.2 is not guaranteed to be true. However, for this to obtain, P and Q must be related in a specific way: it must be the case that some worlds that verify P also verify $\sim Q$, while no worlds that satisfy P also satisfy $\sim Q$. This requires that those worlds which have the same structural profile about P as the actual world would verify $\sim Q$, but that no worlds that have the same structural *and* intrinsic profiles about P as the actual world satisfy $\sim Q$. In this manner, and assuming that the primary and secondary intensions of Q coincide, for P.2 to be false, it must then be the case that the structural profile of physics in the actual world does not necessitate Q, but that the structural and intrinsic profile does (Chalmers, 2010: 151).

Following this reasoning, we reach the crux of our final distinction. As noted in Chapter 2, physics is often characterised by its causal structuralism, such that it defines P by what micro-level entities do, as opposed to what these entities are in-and-of-themselves. In this manner, it does not typically uphold that there is an intrinsic profile to P over that which is described by its primary intensions. In this light, the negative component of Russellian monism was the claim that these micro-level entities must be grounded with an intrinsic nature which *realises* those structures and relations. The positive component of it was its provision of candidate for this: one whose nature is intimately related to conscious experience. Now, as we have seen, for P.2 to be false, it must be the case that the structural properties of P do not necessitate the Q, but that the intrinsic properties of P alongside those structural properties do (ibid: 151). This conclusion perfectly matches the Russellian monist's thesis that the structural properties of

physics in the actual world do not necessitate the phenomenal experience, but that the intrinsic properties of physics plus those structural properties do (ibid: 153). In this manner, we can say that in order to force the falsity of P.2 of our simple Conceivability Argument – the metaphysical possibility $P \& \sim Q$ – the physicalist must concede that some form of Russellian monism is true.

We can now structure our complex Conceivability Argument, with an additional premise. As previously, ‘P’ should be taken to represent the complete pure physical truths about the universe; ‘Q,’ should be taken to represent all phenomenal truths about the universe; ‘&’ should be taken to represent ‘and;’ and ‘ $\sim$ ’ should be taken to represent ‘not.’ Finally, ‘conceivable’ should be taken henceforth – and throughout the remainder of this thesis – to represent ideal, negative, primary conceivability.⁴¹ We can thus formulate our complex Conceivability Argument as follows:

- | | |
|-------|--|
| P.1 | It is conceivable that $P \& \sim Q$. |
| P.2 | If it is conceivable that $P \& \sim Q$, it is primarily possible that $P \& \sim Q$. |
| P.3 | If it is primarily possible that $P \& \sim Q$, then $P \& \sim Q$ is secondarily possible, or Russellian monism is true. |
| P.4 | If it is secondarily possible that $P \& \sim Q$, it is metaphysically possible that $P \& \sim Q$, and so physicalism is false. |
| <hr/> | |
| C. | Physicalism is false or Russellian monism is true. |

(Chalmers, 2010: 161).

As per our simple Conceivability Argument, it will be useful to discuss again exactly what is going on here. First, P.1 has reemphasised the existence of an epistemic gap between P and Q. As noted, if it were the case that P epistemically entailed Q, one could deduce Q from their

⁴¹ Unless qualified otherwise.

knowledge of P. Where the simple Conceivability Argument showed that we are not currently capable of doing this, our complex Conceivability Argument has strengthened this claim to state that no amount of knowledge concerning P will ever sufficiently diminish the epistemic gap between P and Q. Granting this, we could never rule out that $P \& \sim Q$ could obtain. Through qualifying the conceivability of $P \& \sim Q$ as ideal negative conceivability, we extrapolate up to a perspective of ideal reasoning, and so appear to avoid the type-c physicalist objection that we are ignorant of the physical complexities required to decipher claims of necessitation between P and Q. The ideal negative conceivability of $P \& \sim Q$ then amounts to the claim that there will never be entailment between P and Q such that P a priori explains Q, or that Q is deducible from or fully explainable by P.

Now that ideal conceivability has strengthened the claim that there is a lack of epistemic entailment between P and Q, P.2 then acknowledges the primary possibility of $P \& \sim Q$. While this primary possibility has been shown insufficient in-and-of-itself to attack physicalism, the apparent lack of distinct primary and secondary intensions within Q has supported the claim of P.3: that the primary possibility of $P \& \sim Q$ entails the secondary possibility of $P \& \sim Q$, or that Russellian monism is true. By qualifying our notion of conceivability and acknowledging the distinction and relations that hold between the primary and secondary intensions of a term, we have diminished the claim that conceivability is an imperfect guide to possibility: ideal negative primary conceivability is a good guide to primary possibility, and in the case of claims about Q, this primary possibility entails secondary (and so metaphysical) possibility. Since Q lacks distinct primary and secondary intensions, either there is a possible world in which the secondary intension of $P \& \sim Q$ is true, or the secondary intension of $P \& \sim Q$ is false in all possible worlds, but only if the entities in P are furnished with an intrinsic property that is intrinsically relevant to explaining Q.

Finally, since physicalism requires that P constitutively grounds Q, it requires that if P obtains, Q necessarily also obtains in any possible world. Per P.3, $P \& \sim Q$ is either secondarily (and so metaphysically) possible or P must be relevantly furnished with additional properties such that Russellian monism is true. Since the former acknowledges the possibility of a world in which P does not necessitate Q, P must not necessitate Q, and so physicalism must then be false, or Russellian monism must be true. This conclusion is a strong one. P.1 has been strengthened by an ideal perspective about P and Q, and has shown that even from this perspective, there is arguably still a failure of epistemic entailment between P and Q: the Hard Problem is impassable for the physicalist. Granting this epistemic conclusion, one can achieve an ontological conclusion about the natures of P and Q that is either inconsistent with the modal thesis presented by physicalism, or which supports the Russellian monist.

3.2 ~ Defending Physicalism

There are a number of ways in which the physicalist might still respond to the Conceivability Argument. I shall discuss some of the more prevalent objections here alongside their typical rebuttals. The first way is to take issue with the claim of P.1: that $P \& \sim Q$ is conceivable. As discussed, this premise should be understood as invoking ideal primary negative conceivability: that $P \& \sim Q$ cannot be ruled out a priori even from a position of ideal rational reflection. The physicalist might propose that although $P \& \sim Q$ is *prima facie* conceivable, it is not ideally conceivable, and so could yet turn out to be metaphysically impossible (Van Gulick, 1999; Worley, 2003: 15-23). By this reasoning, the epistemic gap that exists between physical and phenomenal truths might still only be a *prima facie* one which could come to be closed as our knowledge of P grows (Chalmers, 2010: 155). For example, new discoveries about physical

processes, neurological compositions, or even the development of some novel reasoning might explain away the epistemic gap.

This form of response harnesses the previous type-c physicalist intuition that we are simply not omniscient about the properties relevant to P. We clearly do not yet know all of what P includes, nor do we know how all of these properties interact. Indeed, few physicists would argue that we currently have a complete physical theory of the universe, or that all of its currently established parts are immune from further amendment or future revolution. Even at the level of neuroscience, there is still clearly much we do not know. And although we might not yet be able to a priori deduce Q from neural-level facts about P, there is still a great deal of progress to be made concerning knowledge about the latter, and a great number of hypotheses regarding it that we have not yet entertained. Granting this valid concern, we might not be in a good position to truly know what P a priori entails, nor then whether P could a priori entail Q. In this manner, upon achieving a greater understanding of P, Q could be explained away. Let us call this objection the **Ignorance Objection**.

The standard response to this form of objection is to acknowledge that although we do not know, nor have we entertained all of the truths about P related to Q, we can know that all such related truths will be of a certain kind. Furthermore, we can also know at this very moment that no truths of that certain kind a priori entail truths about Q. Accordingly, it doesn't matter that we do not yet know all of what P includes, or how all of its relevant properties interact: truths about P appear to be of the wrong kind to ever a priori entail truths about Q. For analogy, it is intuitively the case that no truths about geometry could ever a priori entail positive truths about the smell of a sandwich, and I can intuitively make this claim regardless of whether I know what that sandwich was like in all geometric respects: truths about geometry will always be of

the wrong kind to a priori entail truths about smell (Cutter, 2018). The proponents of Conceivability Arguments who reject the Ignorance Objection can level an analogous charge to P and Q and furnish P.1 with the following background assumptions.

- (i) All truths about P are truths of a particular kind (P_x).
- (ii) No truths of kind P_x a priori entail truths about Q.

There are a few ways in which we could reinforce these assumptions. The first way is to clarify that truths of kind P_x are non-experiential truths, whereas truths about Q are experiential truths. In this manner, we might state the following epistemic gap:

Non-experiential/Experiential Gap: there is no a priori entailment from non-experiential truths to experiential truths (Chalmers, 2016a: 31).

Or more precisely:

For any group of non-experiential parts $P_1, P_2 \dots P_n$ standing in structural relation R, the following scenario is conceivable: there are $P_1, P_2 \dots P_n$ standing in R, but it's not the case that $P_1, P_2 \dots P_n$ standing in R possesses any composite experience.

A different potential refinement would be to choose a value of P_x that corresponds more closely to the subject matter of physics (i.e., not defined in terms of Q) (Cutter, 2019: 111). As discussed in the previous chapters, physics has a mathematico-nomic vocabulary that is only interested in truths about structure and dynamics: it lays down fundamental equations which enable it to deal with the structure of events but does not comment on the intrinsic character of the events that have that structure. We can say that this austere vocabulary includes truths about spatiotemporal, causal/monic, mathematical, and logical expressions – henceforth ‘structural truths’ (Chalmers, 2012). With this in mind, the proponent of Conceivability Arguments who

rejects the Ignorance Objection could propose that truths of the kind P_x are structural truths, and propose the following intuitive epistemic gap:

Structural/Experiential Gap: there is no a priori entailment from structural truths to experiential truths.

Or, more precisely:

For any group of structural properties $P_1, P_2 \dots P_n$ standing in structural relation R , the following scenario is conceivable: there are $P_1, P_2 \dots P_n$ standing in R , but it's not the case that $P_1, P_2 \dots P_n$ standing in R possesses any composite experience.

Whichever epistemic gap one chooses to furnish P.1, the charge remains the same: it is irrelevant that we do not have an ideal knowledge about P at this very moment; all truths about P are of a particular kind, and that kind is not appropriate to deduce facts about Q from. Indeed, it might be intuitively obvious that an experiential truth like 'someone is having the phenomenal experience of smelling coffee' could never be a priori deduced from truths expressed in an austere structural language, or in terms of non-experiential truths. Harnessing either the nonexperiential/experiential gap or the structural/experiential gap, we can arguably abstract to an ideal perspective about these variables without any inconsistency arising. Along these lines, the Ignorance Objection is proposed to be fruitless: regardless of our current ignorance about P , we know that truths about P are of the wrong kind from which to a priori deduce truths about Q . Accordingly, it arguably does not matter what future advancements we make concerning our knowledge about P , or what truths about P a priori entail more broadly; P will never a priori entail Q .

The second way the physicalist might respond to the Conceivability Argument is to accept that although $P \& \sim Q$ might *appear* to be ideally conceivable to us (and indeed, might always), since we are nonideal creatures, we will never be in a position to discern whether or not some given statement is actually ideally conceivable (Bailey, 2007). This objection harnesses the idea that even if we were to have a complete physical theory of the universe, we still might not be able to ideally conceive of $P \& \sim Q$, due to our own cognitive limitations. If this holds, the apparent ideal conceivability of $P \& \sim Q$ as presented by the rebuttal above might arise from our own impaired rationality: if we were only more intelligent we would see that $P \& \sim Q$ is not coherent at all. On this view, $P \& \sim Q$ is not ideally conceivable, and so not logically possible at all – it is just the limitations of our own cognition that misled us into believing so. Accordingly, the argument would be of no threat to physicalist (Chalmers, 1996: 139; Chalmers, 2010: 155). We can call this the **Non-Ideal Objection**.

One might support this line of argument by appealing to analogies regarding the necessity of certain complex mathematical truths that lay beyond human cognitive capacities, viz. certain true mathematical statements that we cannot ourselves prove. Kurt Gödel's first 'incompleteness' theorem states that for any consistent formalized system F , which contains elementary arithmetic, there exists a sentence G_F of the language of the system which is true but unprovable in that system (Raatikainen, 2005: 513). In philosophical parlance, these truths are necessary even though they are not ideally knowable a priori, nor grounded in a combination of priori and empirically knowable factors of the Kripkean form expounded in the previous subsection (Chalmers, 2010: 174; 1996: 139). If our mathematical powers involve nothing more than algorithmic computation, then by Gödel's theorem, there must be certain true mathematical statements that we cannot ourselves prove (Chalmers, 1996: 139). The physicalist might propose that the connection between P and Q is a necessity of the same form

as these truths, which is simply beyond the powers of our cognitive capacities. In this manner, appeals to ideal conceivability about those variables would lack the necessary effectiveness to attack physicalism, as we simply cannot know what is or is not ideally conceivable in the form relevant to render metaphysical possibility: since we are not ideal creatures, it could be the case that $P \& Q$ holds necessarily, but we simply do not have the cognitive capacities to know that truth a priori (ideally or otherwise).

While we are clearly nonideal creatures, one can rebut this response on the grounds that the analogy supporting it is imperfect. In the mathematical case expounded above, our modal reasoning leaves the matter of conceivability open (Chalmers, 1996: 139). While we might be able to conjure some weak sense in which it is conceivable that those complex mathematical truths are false, this does not provide a conceivable world in which they fail to be true. Indeed, ignorance of such truths existing does not equate to the claim that it is conceivable that they do not exist or are false. In the case of P and Q , however, the matter is not left open (ibid). We know of P , and we know of Q . Furthermore, as expounded in the previous rebuttal, we arguably know enough about P to know that truths about it are always of the wrong kind to a priori entail truths about Q . Accordingly, there appears to be a clearly conceivable world in which $P \& \sim Q$. For physicalism to be immune to such a world, one must rule out its possibility despite the best evidence given by our modal powers. Since the mathematical case does not make an analogous case – whereby an apparently possible world is ruled out in this way – it is of no help to the physicalist.

Now, while it may be conceded that any philosophical argument could go wrong because of our situation as nonideal creatures, we have little substantial reason to believe this is the case. Furthermore, in the case of $P \& \sim Q$, we arguably have enough grasp of the kind of truths each

variable delivers to believe that arguments concerning them hold weight. To claim otherwise would appear to be quite *ad hoc* and should be the last resort of the physicalist (ibid: 140). It would relinquish the claim that they have any relevant understanding of the variables their theory is interested in and would further leech into their ability to make any broader philosophical proof. In this manner, it should also be the last resort of philosophers in general: only after we have considered all of the alternatives to physicalism should we turn to the possibility that relation between P and Q is unknowable, or that established philosophical arguments with such variables inputted into them might be mistaken by virtue of our cognitive impairments.

Failing to threaten the claim of P.1, our physicalist might instead turn to those premises arguing for a link between conceivability and possibility. They could, for example, harness the type-b physicalist intuition that although the epistemic gap between P and Q will not be resolved, this is not representative of an ontological problem. Perhaps they could turn to P.2, whereby it was claimed that 'If it is conceivable that $P \& \sim Q$, it is primarily possible that $P \& \sim Q$.' As noted in the previous subsection, this premise invokes the Primary Conceivability Principle, whereby if some sentence X is primarily (ideally negatively) conceivable, then it is primarily possible that X. Or, that if $\sim X$ is not a priori, then X is primarily possible. Now, the type-b physicalist proposes a primitive identity relation between physical and phenomenal states; because it is primitive, it is not deducible even with a complete physical understanding of the world. Through denying this principle on these grounds, the physicalist can claim that while p-zombies are conceivable, they are not primarily possible, and so also not metaphysically possible.

The general link between conceivability and possibility has long been accepted within the history of philosophy and has only more recently been questioned because of the sort of reasoning expounded in the previous subsection, such as the Kripkean case of ‘Water is X_5Y ,’ or ‘Water is not H_2O .’ However, the Primary Conceivability Principle and the Secondary Conceivability Principle have been demonstrated to bypass such examples easily, whilst retaining the link between conceivability and possibility. Accordingly, the onus is on the physicalist to provide a sufficient new counterexample to rival the Primary Conceivability Principle and so P.2. Now, although healthy discussion about such matters exists within the contemporary literature (see Soames (2004), Vaidya (2008: 191-212), and Mallozzi (2021: 1387-1408)), I omit these here for the sake of brevity. On the one hand, such discussions have so far proven inconclusive, and, on the other hand, they simply will not prove relevant to my later analysis within this thesis.

The physicalist might also choose to object to P.3 or P.4. The former claimed that ‘If it is primarily possible that $P \& \sim Q$, then $P \& \sim Q$ is secondarily possible, or Russellian monism is true,’ and the latter claimed that ‘If it is secondarily possible that $P \& \sim Q$, it is metaphysically possible that $P \& \sim Q$, and so physicalism is false.’ To attack the former, the physicalist must provide a clear argument for why the primary and secondary intensions of Q differ, and so make the seemingly nonsensical claim that whatever consciousness feels like is not the essential nature of consciousness. Furthermore, since the converse has already been defended in a vocabulary which is broadly accepted across the community of analytic philosophy the onus is again on the physicalist to prove otherwise (Chalmers, 2004). Similarly, to attack the latter, the physicalist must argue that secondary possibility does not equate to metaphysical possibility, such that the essential nature of an entity does not entail metaphysical truths about that entity. While arguments in this vein do exist – for example, Mizrahi & Morrow (2014) – as before, I omit such argumentation for the sake of brevity and focus.

At this point in my exposition, I have provided little obvious reason to discount the conclusion of our Complex Conceivability Argument against physicalism. Although vast further literature exists regarding such forms of objection and rebuttal, I have kept my focus here brief in light of the particular focus of this thesis. We have considered potential and contemporary objections to each individual premise within the argument, alongside some *prima facie* rebuttals. For example, some objections were rebutted via further epistemic gaps, others were rebutted on the grounds that such forms of objection should be the last resort for any philosopher, and others were rebutted simply on the grounds that the onus is on the physicalist to provide more sufficient evidence to the contrary. With this in mind, the Complex Conceivability Argument appears to currently present a robust challenge against physicalism, and argument for Russellian monism, such that Russellian panpsychism is theoretically preferable (Chalmers, 2010; 1996; Goff, 2017).

3.3 ~ Russellian Panpsychism's Alleged Immunity

In Section 3.1.2, we saw how if the physicalist denies that P does not possess primary and secondary intensions that coincide, then he is led to the conclusion that the structural properties of P do not necessitate the Q, but that the intrinsic properties of P alongside those structural properties do. In this manner, we saw that if the physicalist wishes to force the metaphysical impossibility of $P \& \sim Q$, either physicalism must itself be false, or Russellian monism must be true. While the Complex Conceivability Argument is then both a criticism of physicalism, and argument for Russellian monism, it is also typically an argument that the Russellian monist believes herself to be invulnerable to.

To further explain this alleged invulnerability, it will be useful to distinguish two different ways of defining P. Previously, P has been taken to represent the complete physical truths about the universe, as relevant to physicalism. This was stated to include truths about all of the physical entities, properties, relations, and laws (including those known, and those currently not known but which nonetheless exist throughout space and time) that can be captured via the mathematico-nomic vocabulary of physics. For the sake of clarity here, we should now understand this definition of P to refer to *narrowly physical properties*.

Narrowly physical properties: the structuralist properties traditionally defined by physics in its characterisation of matter (Cutter, 2019: 113)

In addition to this definition of P, we can now introduce a new notion, named *broadly physical properties*. Broadly physical properties should be taken to represent the complete pure physical truths about the universe, as relevant to physicalism, with the addition of the Russellian panpsychist's quiddities.

Broadly physical properties: the structuralist properties traditionally defined by physics in its characterisation of matter plus the categorical qualitative properties which realise those structuralist ones (Cutter, 2019: 113).

With this in mind, we can define the following terms concerning truths about these properties:

Narrowly physical truths: truths concerning the instantiation of narrowly physical properties.

Broadly physical truths: truths concerning the instantiation of broadly physical properties (Cutter, 2019: 113-114).

Similarly, we can define the following terms concerning the grounding of these properties:

Narrow phenomenal grounding: Instantiations of phenomenal properties are grounded in instantiations of narrowly physical properties.

Broad phenomenal grounding: Instantiations of phenomenal properties are grounded in instantiations of broadly physical properties (Cutter, 2018: 48).

Although these definitions may be contentious to some, the clear delineation they help make between our rivaling theses is of great help when utilised within the purview of Conceivability Arguments. Accordingly, and with these clarifications in place, we can say that physicalism entails the thesis that phenomenal experience is grounded solely in narrowly physical truths. Conversely, we can say that Russellian panpsychism entails the thesis that phenomenal experience is grounded in broadly phenomenal truths, and so not wholly narrowly physical truths.

Utilising these definitions, we can illuminate the why the Russellian panpsychist typically views their theory to be immune to the Conceivability Argument. When we are requested to conceive of p-zombies, or that $P \& \sim Q$, we are requested to conceive of an entity which is identical to us in all physical respects, but which differs from us in experiential respects (in this instance, it has no qualitative experience). Now, when one attempts to conceive of this, they harness facts about all the physical features we understand ourselves to possess. Presumably, those physical properties that one holds fixed are those which are described by the traditional physical sciences. We can say then that when we conceive of $P \& \sim Q$ in our traditional Conceivability Argument, we define P in terms of narrowly physical properties – the physical properties traditionally defined by physics.

For the Russellian panpsychist, however, P should be defined in terms of broadly physical properties. Since broadly physical properties include quiddities – and so micro-experiential instantiations of phenomenal experience – P has Q built into it already. Since P has Q built into it, however, conceiving of P without Q is arguably impossible: it allegedly cannot be the case that P (when P refers to broadly physical properties) and not Q. In this manner, the Russellian panpsychist is allegedly immune to Conceivability Arguments about those variables, as $P \& \sim Q$, when appropriately defined becomes inconceivable. It follows that for the Russellian panpsychist, $P \& \sim Q$ is inconsistent, as if P (when defined as broadly physical properties), then Q, necessarily. (Cutter, 2019; Stoljar, 2001; Alter & Nagasawa, 2012; Chalmers, 2016a).

3.4 ~ Concluding Remarks

This chapter has introduced one of the popular arguments utilised by Russellian panpsychists to illustrate their theories alleged theoretical advantage over physicalism: the Conceivability Argument. After introducing the Conceivability Argument in its ‘simple’ form, I delved into Chalmers’ more complex 2-Dimensional variant, where he hones a particularly powerful definition of conceivability, establishes a link between this conceivability and possibility, and concludes that claims about metaphysical possibility can be validly made via a priori reasoning. After delving into Chalmers’ explicit argument against the physicalist in this regard, I showed how the proponent of this Conceivability Argument is led to the conclusion that either physicalism is false, or Russellian monism is true. If Chalmers’ argument is sound within this vein, then the Russellian panpsychism possesses an argument for its theoretical superiority over physicalism.

I then discussed a number of ways in which the physicalist could respond to the Conceivability Argument against them. I considered appeals to ignorance, alongside attempts to deny

Chalmers' explicit links between conceivability and possibility. While the former objections were negated via further epistemic gaps, the latter necessitate the physicalist to provide new counterexamples and counter-principles to these well-established links. After establishing the strength of Conceivability Arguments against physicalism, I then analysed the typical Russellian panpsychist claim that their theory is invulnerable to them. Through distinguishing two definitions of 'physical properties,' I explained how the statement $P \& \sim Q$ is inconceivable to the Russellian panpsychist, when appropriately defined: if P (when defined via broadly physical properties), then Q, necessarily; and, if P then Q, then Russellian panpsychism is immune to the Conceivability Argument. At this point, the Conceivability Argument appears to present a robust challenge against the physicalist: through a priori means, it appears to establish that either physicalism is false, or Russellian monism is true. Furthermore, since Russellian monism appears immune to such forms of argumentation, it appears to clearly have the upper ground against the physicalist. *ceteris paribus*, physicalism is false, and Russellian monism is true.

4

The Misuse of Conceivability Arguments

Applications and Contradictions

The Conceivability Argument is typically levelled as a robust argument against physicalism, and argument for Russellian panpsychism. However, I shall acknowledge how the Combination Problem can be harnessed to produce an analogous argument and conclusion against Russellian panpsychism. In this manner, I will argue that those who uphold Conceivability Arguments in their motivation for Russellian panpsychism will find that motivation undermined. However, I shall ultimately propose that the application of Conceivability Arguments to the Mind-Body debate is flawed, and so they should not be appealed to at all. This contentious move removes one of the most popularly appealed to arguments for Russellian monism from their arsenal, and so necessitates a new and improved argument to distinguish it from its rivals.

In order to show the flaw within Conceivability Arguments, I will assume throughout that the argument is itself valid. In Section 4.1, I will build from Brian Cutter (2019) and Philip Goff (2009) to produce a Conceivability Argument against Russellian panpsychism which carries equivalent weight to its physicalist counterpart. After considering some potential – but ultimately unsuccessful – rebuttals to this argument in Section 4.2, I will conclude that Russellian panpsychists who uphold Conceivability Arguments in support of their theory will find their theory undermined.

In Section 4.3, I shall then argue that although Russellian panpsychism is vulnerable to Conceivability Arguments, *all* philosophy of mind hypotheses are. While some authors have already levelled Conceivability Arguments to philosophy of mind hypotheses other than physicalism (e.g. Goff, 2017; Chalmers, 2016b; Aleksiev, 2023; Brown, 2010; Cutter, 2019), none have yet acknowledged that they can be levelled to *all* hypotheses, nor then explored the potential consequences of this extension; I shall bring all of these arguments together for the first time, and explore their connections and potential consequences. In this light, and after proposing Conceivability Arguments against Russellian panprotopsyism, substance dualism, and property dualism, I shall argue that – assuming each argument has equal force – there are three clear possibilities that one may draw. Either: (i) each of my proposed arguments’ premises are true in each application, and that our entire spectrum of hypotheses are incorrect; or, (ii) at least one of their premises is false in each application, and so they cannot provide us with a true conclusion about any philosophy of mind hypothesis; or, (iii) an adequate justification for the truth of at least one of their premises is currently unattainable, and so they are currently of no use within the Mind-Body debate.⁴²

In Section 4.4, I will construct a series of valid Conceivability Arguments which negate a metaphysical truth about the world from which they are posited. Building from Katalin Balog’s criticism of Conceivability Arguments (1999), alongside Chalmers’ subsequent rebuttal (2010), I shall construct three ‘Counterfactual Conceivability Arguments’ which bypass Chalmers’ concerns and establish Balog’s intended conclusion: that there is a flawed premise within Conceivability Arguments. This analysis will show that the vulnerability of our philosophy of mind hypotheses to Conceivability Arguments is not necessarily a fault of those hypotheses

⁴² It must be emphasised that these conclusions only apply to the Conceivability Arguments within this thesis (with their application solely to the consciousness debate); I do not intend for these conclusions to be extended to all conceivability arguments, *simpliciter*.

themselves, but rather a fault with the argument. In this manner, I shall discount consequence (i). I shall then locate this fault primarily in P.1 of the Conceivability Argument, specifically with regards to the claim that any of the stated scenarios are ‘ideally’ conceivable. I will then locate a further issue in the Chalmers’ specific claims about the ‘primary’ and ‘secondary’ intensions of Q. Finally, I will harness these sub conclusions to propose that (iii) is the most likely.

I shall conclude that those who wish to propose Conceivability Arguments as an argument against physicalism will find their own theory vulnerable. However, since I show this to be a fault of Conceivability Arguments, as opposed to a fault within any philosophy of mind hypothesis, I recommend that they are relinquished from the Russellian panpsychist’s repertoire. Doing so removes one the most popularly appealed to arguments for the theory, but such a move is pertinent if the Russellian panpsychist wishes to make a compelling case. This restriction of Conceivability Arguments will then reinforce my new and improved argument for Russellian panpsychism in the subsequent chapter.

4.1 ~ Conceivability Arguments against Russellian Panpsychism

This section will assess and deny the claim that Conceivability Arguments are beneficial to Russellian panpsychism. Although the Complex Conceivability Argument appears to establish that either physicalism is false or Russellian monism is true, I will show how a similar Conceivability Argument can be used to show the alleged falsity of Russellian panpsychism.⁴³ Building from Philip Goff (2009) and Brian Cutter (2019), I will show how the Combination Problem for Russellian panpsychism can be harnessed into a number of analogous epistemic

⁴³ In 4.3, I will show how a similar Conceivability Argument can also be levelled to Russellian panprotopsyichism.

gaps to those formed out of physicalism's Hard Problem. These epistemic gaps will then be inputted into a new Conceivability Argument which – by the same logic as the original argument – show Russellian panpsychism to be false. In this manner, I will show that Conceivability Arguments do not actually establish Russellian panpsychism's theoretical benefit over physicalism, as Russellian panpsychism itself remains susceptible to them.

We have seen that the physicalist's vulnerability to the Conceivability Argument comes from the apparent epistemic gap that exists between narrowly physical truths and phenomenal truths. Initially, this was formulated from the epistemic issue illustrated by the Hard Problem and was then reinforced through the non-experiential/experiential gap, and the structural/experiential gap. In this manner, we can say that the Conceivability Argument against physicalism relies on a general epistemic gap between P and Q.

In Section 2.3.2, I introduced William James's Combination Problem: The Subject Summing Problem (James, 1890). Here, it was commented that the property of experience seems to require a subject as a prerequisite: any instance of experience appears to necessitate that it be realised by something, and that something must intuitively be an 'experiencer' in the sense that the experience is 'for it.' Since quiddities are proposed to be phenomenal properties, quiddities must then be realised by micro-subjects. Per Russellian panpsychism, these phenomenal properties – when appropriately arranged – are supposed to constitute our macro-phenomenal properties, and so micro-subjects are supposed to constitute macro-subjects. However, as the Subject Summing Problem illustrated, there appears to be an issue in the idea of micro-subjects combining to make a further subject (such as a macro-subject).

With the Combination Problem in mind, Goff introduces the notion of ‘panpsychist zombies’ or, ‘P⁺-zombies’ (Goff, 2009: 296). These are creatures that are broadly physically identical to us (in terms of the structural properties and micro-experiential quiddities of their parts), but which lack consciousness as a collective whole. To illustrate this, let us conceive of a traditional p-zombie (in all the narrowly physical respects). Next, we should attempt to ‘colour in’ all of its narrowly physical parts with quiddities, and the relevant phenomenal properties associated to them. Goff proposes that no matter how we colour in the parts of our zombie, however, it remains conceivable that no further macro-subject emerges (ibid). In this manner, we could propose that it is conceivable that P⁺ (where P⁺ refers to all of the broadly physical truths about the universe) but not Q (where Q refers to some collective macro-subject).

There are a number of ways in which we can formulate epistemic gaps to furnish the notion of P⁺-zombies. Brian Cutter, for example, focusses upon James’ comment that a ‘group’ consciousness could never be deduced from the subject-parts said to constitute it, nor could it be said to have evolved from those parts (Cutter, 2019: 114-115; James, 1890: 160). In this vein, our Russellian panpsychist might understand that the macro-experience of severe pain arises from millions of micro-experiences of pain instantiated across one’s neurons. However, one still cannot deduce the any composite macro-experience from those micro-experiential neurons, as there is no deductive inference one can make of the form ‘micro-experiences → macro-experience’ (Goff, 2017:173-175). Building from this, Cutter claims that the Russellian panpsychism is burdened with the issue that no set of subjects seems to necessitate a further subject, since no composite subject could ever be a priori deduced from subject-parts (Cutter, 2019: 115). Furthermore, since micro-experiences don’t necessitate any macro-experience, we can conceive of the former in the absence of the latter. To illustrate this issue more precisely, we can formulate the following epistemic gap, built from both Cutter and Goff:

Subject/Subject Gap: For any group of subjects, $S_1, S_2 \dots S_n$ which instantiate phenomenal properties, $Q_1, Q_2 \dots Q_n$ and stand in structural relation R , the following scenario is conceivable: there are $S_1, S_2 \dots S_n$ instantiating $Q_1, Q_2 \dots Q_n$ and standing in R , but it's not the case that there is a subject S_{macro} such that S_{macro} is not identical with any of $S_1, S_2 \dots S_n$ (Cutter, 2019: 116; Goff, 2017: 174; Goff, 2009).

Goff provides a negative example to aid with the comprehension of this sort of gap. When we think of a party, we think of a group of people dancing, drinking, and acting merry. Goff argues that it is impossible for one to conceive of a group of people behaving in this way without also conceiving of a party; the two states of affairs cannot intelligibly be separated (Goff, 2017: 174). Similarly, it is arguably impossible to conceive of a state of affairs in which iron plates and beams are arranged and directed identically to how they are in the Eiffel Tower, without also conceiving of a tall structure.⁴⁴ In both of these examples, there is no epistemic gap between the two respective states of affairs. In contrast, when we conceive of two states of affairs involving distinct subjects, those states of affairs *can* be intelligibly separated. As Goff writes, “either can be conceived of obtaining in the absence of another” (ibid: 174). In this manner, it is not the case that when we conceive of a group of subjects $S_1, S_2 \dots S_n$ that we must also conceive of some further subject or macro-subject.

Cutter proposes a second potential epistemic gap faced by Russellian panpsychism which builds from the Subject/Subject Gap (Cutter, 2019: 116-118). To illustrate this, let us attempt to conceive of a narrowly physical isomorph to ourselves. Given the structural/experiential gap of physicalism (that there is no a priori entailment from structural truths to experiential truths),

⁴⁴ Unless, of course, one supports mereological nihilism, whereby one believes that no composite objects exist (e.g. Unger (1979; 1990) and Sider (2001; 2013)).

we are unable to a priori deduce any experiential properties of the isomorph: it could be a p-zombie. Now, let us attempt to colour in all of the isomorphs narrowly physical parts with truths about their alleged phenomenal properties (in such a manner that we add broadly physical properties to the picture). Given the Subject/Subject Gap, we will be unable to a priori deduce any macro-experiential properties of the isomorph: it could be a panpsychist-zombie (ibid: 117). Importantly, there does not seem to be anything special about the fact that we added information about the phenomenal properties of the isomorph's component parts: no information about the intrinsic nature of its parts appears to a priori entail truths about the macro-experience of the whole (ibid). In this manner, there appears to be a clear sense in which there is also a general epistemic gap between 'micro-level' information about the system, and the 'macro-level' experiential information associated with that system as a whole. As with our subject/subject gap, the former state of affairs can intelligibly be separated from the latter state of affairs, and so one can conceive of the former without necessarily conceiving of the latter.

With this in mind, we can formulate our second potential epistemic gap for Russellian panpsychism: a Micro-level/Macro-level Gap. 'Micro-level' should be taken to refer to truths about the intrinsic phenomenal nature of the component parts and their structural relations, and 'Macro-level' should be taken to refer to truths about the phenomenal nature of the system as a whole:

Micro-level/Macro-level Gap: For any group of parts, $P_1, P_2...P_n$ which instantiate phenomenal properties, $Q_1, Q_2...Q_n$ and stand in structural relation R , the following scenario is conceivable: there are $P_1, P_2...P_n$ instantiating $Q_1, Q_2...Q_n$ and standing in R , but where nothing of which $P_1, P_2...P_n$ are proper parts is conscious.

We can use either the Subject/Subject Gap, or the Micro-level/Macro-level Gap to produce a Conceivability Argument against Russellian panpsychism.⁴⁵ I shall set out the argument below, explain how Goff and Cutter’s epistemic gaps can be used to support P.1, and then defend the argument in its conclusion. ‘P⁺’ should now be taken to represent the complete broadly physical truths about the universe, and ‘Q,’ should be taken to represent macro phenomenal truths about the universe.

P.1	It is conceivable that P ⁺ &~Q.
P.2	If it is conceivable that P ⁺ &~Q, it is primarily possible that P ⁺ &~Q.
P.3	If it is primarily possible that P ⁺ &~Q, then P ⁺ &~Q is secondarily possible.
P.4	If it is secondarily possible that P ⁺ &~Q, it is metaphysically possible that P ⁺ &~Q, and so Russellian panpsychism is false.
C.	Russellian panpsychism is false. ⁴⁶

(Goff, 2009: 297).

Both the Subject/Subject Gap and the Micro-level/Macro-level Gap can be used to defend the claim that P⁺&~Q is conceivable, as both gaps entail that the macro-level states of affairs can intelligibly be separated from their micro-level states of affairs. In the case of both gaps, we can appeal to the aforementioned P⁺-zombies. For the former, we can conceive of a broadly physical isomorph – as described in terms of narrowly physical properties, plus quiddities (and including the notion of micro-subjects and the relevant phenomenal properties) – but where there is no further subject entailed at any other level. For the latter, we can conceive of a broadly physical isomorph – as described in terms of the intrinsic phenomenal nature of the isomorph’s

⁴⁵ To aid one’s conception, we can refer back to Stoljar’s race of tiny conscious aliens (2006), or, to Block’s Chinese Nation against physicalism (1980). Since the Chinese Nation is already formed of conscious inhabitants, we can use the exact same thought experiment – without modification – to reinforce these gaps against Russellian panpsychism. In this manner, it might work even better as an intuition pump against Russellian panpsychism than it did against physicalism.

⁴⁶ Cf. Goff (2017), Chalmers (2016b), and Aleksiev (2023) for similar lines of argumentation.

macro-level parts, alongside their structural relations – but where the isomorph has no macro-level experience.

For these conceived scenarios to truly threaten Russellian panpsychism, our notion of conceivability must remain one of ideal, negative, primary conceivability. In the case of our P^+ -zombies, this requires that even from a perspective of ideal reasoning, no amount of a priori reasoning could reveal that $P^+ \& \sim Q$ is incoherent: the epistemic gap (as defined in terms of the Subject/Subject Gap or the Micro-level/Macro-level Gap) must remain impassable even with an idealised knowledge of P^+ and Q . I believe that one can satisfy this notion of conceivability about $P^+ \& \sim Q$ in the same way it was satisfied against physicalism.⁴⁷ It was proposed that as long as the Hard Problem remains for physicalism, there is an epistemic gap between P and Q by definition. Furthermore, it was proposed that the Hard Problem might actually be theoretically impassable, and so that $P \& \sim Q$ would hold even upon ideal reflection. In this same manner, we can say that as long as the Combination Problem remains for Russellian panpsychism, the epistemic gap between P^+ and Q remains. If, for whatever case, the Combination Problem is also theoretically impassable, then $P^+ \& \sim Q$ would hold even upon ideal reflection.⁴⁸

Since it is conceivable that $P^+ \& \sim Q$, it is primarily possible that $P^+ \& \sim Q$ (P.2). Like our Complex Conceivability Argument, this premise upholds the Primary Conceivability Principle: that if some sentence X is primarily (ideally negatively) conceivable, then it is primarily possible that

⁴⁷ One may feel there is more potential to solve the Combination Problem than the Hard Problem, due to the former enjoying the same kinds of concept on either side of the gap. While this may be true, the potential for one to solve either gap has no bearing on current claims about what is conceivable: since epistemic entailment doesn't come in degrees, as long as the epistemic gap between P^+ and Q remains, $P^+ \& \sim Q$ remains conceivable.

⁴⁸ There are a number of potential ways in which the Russellian panpsychist might respond to this argument. These will be considered shortly. As we will see, they fall vulnerable to the same criticisms levelled to the physicalist's responses in Section 3.2.

X. Now, one might claim that the link from primary to secondary possibility here is not as clear as it was in the case against physicalism. Indeed, if the physicalist retained causal structuralism, the primary and secondary intensions of P would coincide. His only option to resist this was to claim a different intrinsic profile of P, which in turn vindicated the Russellian monist. However, while this separated the primary and secondary intensions of P for the physicalist, the case is a little different for the Russellian panpsychist. Since the Russellian panpsychist has redefined P in terms of P^+ (broadly physical properties), she has furnished P with intrinsic profile Q. For the Russellian panpsychist, the thing that plays the mass role, is a now thing for which there is something *that it is like* to be it. Furthermore, since the something that it is like to be a broadly physical property is to realise those particular roles, there is a sense in which the appearance and essential properties of P^+ cannot be distinguished from one another (once those properties are fully understood).

Continuing to uphold Chalmers' interpretation of Kripke's view – that terms expressing conscious states lack distinct primary and secondary intensions – we can say then that since we have established the primary possibility of $P^+ \& \sim Q$, we can infer that the secondary intension of $P^+ \& \sim Q$ is possibly true, and so that it is secondarily possible (P.3 – P.4).⁴⁹ The only option to resist this would be to propose that the intrinsic nature of P does not explain Q, but this would remove the positive component of Russellian panpsychism described in Chapter 2. Finally, since Russellian panpsychism proposes that P^+ constitutively grounds Q, it requires that if P^+ obtains, Q obtains, necessarily. However, since P.3 acknowledges the possibility of a world in which $P^+ \& \sim Q$, Russellian panpsychism must be false (P.4) (Cutter, 2019; Goff, 2009). In this manner, Russellian panpsychists who support Conceivability Arguments against

⁴⁹ In some ways, this link is even stronger than it was against the physicalist, as P^+ and Q now both share the same *type* of essential property.

physicalism without providing a resolve to the Combination Problem might find their theory undermined.

4.2 ~ Defending Russellian Panpsychism

Like physicalism, there are a number of ways in which the Russellian panpsychist might respond to the Conceivability Argument. The first way is to take issue with the claim of P1: that $P^+ \& \sim Q$ is conceivable. As discussed, this premise should be understood as invoking ideal primary negative conceivability: that $P^+ \& \sim Q$ cannot be ruled out a priori even from a position of ideal rational reflection. The Russellian panpsychist might propose that although $P^+ \& \sim Q$ appears *prima facie* conceivable, it is not ideally conceivable, and so could yet turn out to be metaphysically impossible. For this to obtain, it must be the case that the both the Subject/Subject Gap and the Micro-level/Macro-level Gap only present *prima facie* epistemic issues, which could come to be closed on ideal reflection. For example, new theoretical promise into how quiddities can combine to create macro-subjects might explain away the epistemic gaps, such that $P^+ \& \sim Q$ is rendered inconceivable.

Like the physicalist, this form of response harnesses the intuition that we are simply not omniscient about properties P^+ . For example, it might be said that we do not know what experience or subjecthood could be like at the micro-level, nor how those entities might possibly interact, and so we cannot be fairly tasked to determine what is a priori deducible from them. Indeed, since Russellian panpsychism is not widely accepted, nor has it enjoyed a long history of research, it is fair to say that there is clearly much we do not know (regardless of whether or not it turns out to be correct). Accordingly, although we might not yet be able to a priori deduce Q from theoretical facts about P^+ , there is still a great deal of progress to be made concerning knowledge about the latter, and a great number of hypotheses concerning it that we

have not yet entertained.⁵⁰ Like physicalism then, the Russellian panpsychist might propose that we are simply not in a good position to truly know a priori what P^+ entails.

I do not believe that this appeal to ignorance should possess any greater dialectical force than the physicalist's Ignorance Objection. Like in the case of physicalism, we could reinforce P.1 with more detail to our background assumptions, such as:

- (i) All truths about P^+ necessarily include truths about subjects.
- (ii) No truths about any collection of subjects a priori entail truths about a composite subject that isn't already in that collection of subjects.

Or perhaps:

- (i) All truths about P^+ are of a particular kind (micro-level experience).
- (ii) No micro-level information a priori entails truths about macro-level experience.

The former set of background assumption reemphasises the Subject/Subject Gap in such a way that $P^+ \& \sim Q$ should hold ideally, regardless of the fact that we do not have an ideal knowledge of P^+ at this very moment. The latter set of background assumption harnesses the intuition of the Micro-level/Macro-level gap to propose that our experiential properties at the macro-level are inferentially insulated from micro-level information.⁵¹ Whichever background assumption one chooses to furnish P.1 with, the charge to the Russellian panpsychist remains analogous to the charge to the physicalist: just as all truths about P are of a particular kind, and that kind is not appropriate to deduce facts about Q from, so too are all truths about P^+ of a particular kind,

⁵⁰ E.g. those proposed by Roelofs (2019), Miller (2018), Goff (2016), and Mørch (2014).

⁵¹ This is not to say that all macro-level information is inferentially insulated from low-level information. Indeed, much macro-level information is a priori derivable from micro-level information, e.g., the shape of a composite object. Rather, this is just to say that macro-level experiential properties are insulated from information regarding micro-level experience.

and that kind is not appropriate from which to deduce facts about Q. In this manner, an appeal to ignorance from the Russellian panpsychist possesses no greater dialectical force than the physicalist's attempt.

In fact, it could be said that the Russellian panpsychist's appeal to ignorance should possess even *less* dialectical force than the physicalists. The vocabulary adopted by physicalism has enjoyed a rich and rewarding history: many of the mysteries of the universe have been explained away using solely P (narrowly) – and indeed, continue to be explained away. In this manner, it would appear reasonable to place some faith in it explaining away its epistemic gap. Russellian panpsychism, on the other hand, has not enjoyed such a history. We have not shown the existence of quiddities, nor is there a widely accepted form of integration that they may come to enjoy within contemporary science.⁵² In this manner, one might question why we should entertain such an appeal to ignorance at all. Quiddities were proposed to provide the categorical properties which underlie and characterise the micro-level entities posited by physics, and which also transparently explain conscious experience. However, if they falter from epistemic gaps so early – especially since those gaps cast shadow on the claim quiddities can transparently explain (macro) conscious experience – then we might have little motivation to suppose that such gaps will come to be explained away as the theory develops.⁵³

A different way the Russellian panpsychist might respond to the Conceivability Argument is to accept that although $P^+ \& \sim Q$ might appear ideally conceivable, since we are nonideal creatures, we will never be in a position to discern whether or not the scenario is actually ideally

⁵² While one might allege that less work has gone into the Russellian panpsychist project, since we did not afford appeals to ignorance by the physicalist, we should not here either.

⁵³ While potential resolutions have been proposed – for example, by Roelofs (2019), Miller (2018), Goff (2016), and Mørch (2014) – aside from the forthcoming description of Roelofs, I have mostly omitted these here for the sake of fairness in my analysis against physicalism.

conceivable. Like the physicalist, the Russellian panpsychist might propose that the apparent ideal conceivability of $P^+ \& \sim Q$ arises from our own impaired rationality. If this obtains, $P^+ \& \sim Q$ might be neither ideally conceivable, nor logically possible, and would therefore be of no threat to Russellian panpsychism. Now, while the Russellian panpsychist *could* take this line of response, it is again unclear why it should possess any greater dialectical force than that previously given by the physicalist. Again, her reasoning mirrors the physicalist's, and where the physicalist faltered, so too will the Russellian panpsychist. Furthermore, one might again propose that to grant the Russellian panpsychist any benefit of the doubt on such matters would be ad hoc and unwarranted, and that if we are prepared to do such, we must also be prepared to furnish the physicalist's response with an analogous – or perhaps even greater – benefit of the doubt.

Failing to establish any adequate response against the claim that it is conceivable that $P^+ \& \sim Q$, the Russellian panpsychist might then argue against the link between conceivability and possibility. Like the physicalist, they could turn to P.2, where it claimed that 'If it is conceivable that $P^+ \& \sim Q$, it is primarily possible that $P^+ \& \sim Q$.' Or, similarly, she could reject the claim of P.3, that 'If it is primarily possible that $P^+ \& \sim Q$, then $P^+ \& \sim Q$ is secondarily possible,' or P.4, that 'If it is secondarily possible that $P^+ \& \sim Q$, it is metaphysically possible that $P^+ \& \sim Q$, and so Russellian panpsychism is false.' As noted, I believe that it is unclear why one would make such a move.⁵⁴ Furthermore, since she already denied the physicalist this move, to make it herself would be to free physicalism of its own threat: if the Russellian panpsychist rejects any of P.2-P.4 of the Conceivability Argument against her, she rejects her own Complex Conceivability Argument against the physicalist.

⁵⁴ Noted on page 110

There are other attempted responses to these forms of argument, and indeed, other responses to the epistemic gaps that Russellian panpsychism allegedly suffers. For example, with regards to the Subject/Subject Gap, Luke Roelofs proposes that there are some set of conditions that component subjects can meet, such that it will become inconceivable that the whole they form is not itself a conscious subject. They propose to show that such entailments can in fact be shown, such that facts about subjects and their experiences and relations conceptually entails that the whole they form is also a subject of (some of) those same experiences (Roelofs, 2019: 55, 63-65). While they also entertain that it need not be the case that micro-subjects explaining macro-subjects requires that a sufficiently ideal reasoner could see, from the conception of micro-subjects, that macro-subjects follow necessarily, they set aside this latter response on the grounds that it relaxes the standards for counting as a good explanation (ibid, 2019: 61). For example, they propose that if one knows all the microphysical facts about how water molecules interact at room temperature alongside understanding what it means to be ‘liquid,’ the conclusion that water is liquid at room temperature follows entirely a priori, and the negation of this seems contradictory and inconceivable. Since this seems to be what makes physical combination so explanatory, it is a reasonable standard to hope experiential combination can meet also (ibid).

While Roelofs makes a compelling case what is required for composite subjectivity, and posits a potential solution, I shall not delve into any further detail regarding it at this point.⁵⁵ My reasoning is grounded in the fact that it will have no bearing on the particular focus and line of criticism that I take in this chapter: criticising Conceivability Arguments simpliciter. Furthermore, I wish for my dialectic to be fair against each philosophy of mind hypotheses I

⁵⁵ Their explicit proposal will be delved into in more detail in the final chapter.

assess against Conceivability Arguments; there are also many responses to the Hard Problem that I have not considered in this thesis.

We have seen that the Conceivability Argument against Russellian panpsychism can be just as threatening to Russellian panpsychism as the Complex Conceivability Argument was to physicalism. Each of the available lines of response have been shown to possess the same dialectical force as the physicalist's (and, in some cases, perhaps even less). Accordingly, the grounds for rejecting physicalism by virtue of the Conceivability Argument can be extended to Russellian panpsychism. Now, this is not to say that Russellian panpsychism is necessarily false; as it will be shown later in this chapter, the fault lies with our application of the argument, and not necessarily within the theories themselves. However, it does mean that those Russellian panpsychists who reject physicalism on the basis of the Conceivability Argument find the motivation for their own theory undermined. The reality of this charge is strong: it does not require that the Subject/Subject Gap or the Micro-level/Macro-level Gap must be impassable. Rather, it simply requires that such gaps possess at least as much dialectical force as those arguments used to motivate P.1 of the Conceivability Argument against physicalism. As I have argued, at present, they have at least as much epistemic force, and so the Russellian panpsychist must either reject Conceivability Arguments simpliciter or accept that she suffers them to an approximately equal level to the physicalist. Either way, the Russellian panpsychist who attempts to motivate her theory via Conceivability Arguments appears to be mistaken.

4.3 ~ The Universal Applicability of Conceivability Arguments

Conceivability Arguments appear to be equally threatening to Russellian panpsychism. Because of this, they cannot be used to help motivate Russellian panpsychism over physicalism. In fact, if we are to put out faith in them, they might undermine the motivation

for Russellian panpsychism entirely. However, I intend to illustrate a flaw within Conceivability Arguments and recommend that we should dispel them from the debate entirely. To help illustrate this flaw, this subsection will introduce the charge that they can be validly applied to any philosophy of mind hypothesis.⁵⁶ In particular, I will build from Brian Cutter (2019) and Richard Brown's (2010) proposals that one can also construct epistemic gaps into valid Conceivability Arguments against both Russellian panprotopsychism and dualism. Crucially, since epistemic entailment does not come in degrees, it does not matter if any one of the particular epistemic gaps motivating these Conceivability Arguments seems intuitively stronger or weaker than another. If any epistemic gap has not yet been explained away, then there is no current epistemic entailment. Accordingly, such gaps can be harnessed into analogous Conceivability Arguments which should each be afforded the same strength. Furthermore, whichever gap one feels is most strong currently might simply be a result of their particular antecedent assumptions, and so discarding others on the hope that they will be explained away would be premature.

As noted previously, while many authors have posited Conceivability Arguments against a wide range of philosophy of mind hypotheses, none have yet entertained the potential consequences of this breadth of application. Granting that each of my proposed arguments possess equal dialectical force and assuming – as I continue to do – that the basic argument is valid, I will propose that there are three clear possibilities that one may draw from my analysis. Either:

- (i) All of the arguments' premises are true in each application, and our entire spectrum of hypotheses (physicalism, dualism, and Russellian monism) are incorrect. Or,

⁵⁶ The reasons for this will be addressed in Section 4.4.

- (ii) At least one of their premises is false in each application, and so they cannot provide us with a true conclusion about any philosophy of mind hypothesis. Or,
- (iii) An adequate justification for the truth of at least one of their premises in each application is currently unattainable, and so they are currently of no use within the Mind-Body debate.

While each conclusion establishes that Conceivability Arguments pose no more threat to Russellian panpsychism than any other theory, whichever conclusion one takes, the case for Conceivability Arguments being of any benefit to Russellian panpsychism seems to be diminished further. Furthermore, if (ii) or (iii) are correct, then Conceivability Arguments are of no use to the Mind-Body debate and should be relinquished entirely.

Brian Cutter (2019) motivates a Conceivability Argument against (constitutive and microprotopsychist) Russellian panprotopsychism via similar epistemic gaps to those faced by Russellian panpsychism and physicalism. Recall, Russellian panprotopsychism is the thesis that protophenomenal properties underlie and characterise the micro-level entities posited by physics in its characterisation of matter, and are intrinsically suited to bringing about phenomenal properties, when under the appropriate conditions. In this manner, protophenomenal properties were defined as properties which are not themselves phenomenal, but which hold some exceptionally close relationship to phenomenal properties, such that they ground conscious experience in the same way that basic physical properties ground non-basic properties such as temperature.

Similar to the case against Russellian panpsychism, we can introduce the notion of ‘panprotopsychist zombies’ or, ‘P*-zombies.’ This is a creature that is broadly physically

identical to me (in terms of structural properties and protophenomenal quiddities), but which lacks consciousness as a collective whole. To illustrate this, let us again attempt to conceive of a traditional p-zombie (in all the narrowly physical respects). Next, we should attempt to ‘colour in’ all of its narrowly physical parts with quiddities, and the relevant protophenomenal properties associated to them. No matter how we colour in the parts of our zombie, however, it remains conceivable that no further macro-subject emerges. In this manner, we could propose that it is conceivable that P* (where P* refers to all of the broadly physical truths about the universe) but not Q (where Q refers to some collective macro-subject).

Cutter proposes a number of ways in which we can formulate epistemic gaps to furnish the notion of P*-zombies. The first way we can do this is via a similar gap to the non-experiential/experiential gap (Cutter, 2019: 116). For physicalism, this gap harnessed the intuition that all truths about P are of a particular kind (non-experiential), and that no truths of that kind a priori entail truths about Q. Or, more precisely, that there is no a priori entailment from non-experiential truths to experiential truths (Chalmers, 2016a: 31). Now, by defining protophenomenal properties as properties which are not themselves phenomenal – but which constitute conscious experience in the same way that basic physical properties constitute non-basic ones – the Russellian panprotopsychoist immediately makes herself susceptible to a similar gap:

Protophenomenal/Phenomenal Gap: there is no a priori entailment from protophenomenal truths to phenomenal truths.⁵⁷

⁵⁷ As before, one might think the similar epistemic gap against physicalism is still stronger, since panprotopsychoism enjoys *similar* kinds of concept on either side of the gap. While this may be true, the potential for one to solve either gap has no bearing on current claims about what is conceivable, since epistemic entailment doesn’t come in degrees.

This gap can be defended via analogous reasoning to the defence of the non-experiential/experiential gap and carries an equal epistemic threat: just like how no amount of geometrical truths about a sandwich would ever a priori entail truths about how it smells, no amount of knowledge about properties which are not themselves phenomenal will ever a priori entail truths about phenomenal properties; the former is simply of the wrong kind to ever a priori entail truths about the latter. To reinforce this claim further, we can redefine this gap as follows:

For any group of physical parts with proto-phenomenal properties $P_1, P_2 \dots P_n$ standing in structural relation R , the following scenario is conceivable: there are $P_1, P_2 \dots P_n$ standing in R , but it's not the case that $P_1, P_2 \dots P_n$ standing in R as a composite system is conscious.

Cutter proposes that we can motivate a second epistemic gap against Russellian panprotopsychism via the previously discussed Micro-level/Macro-level gap, and for the same reasons (Cutter, 2019: 118). This time, 'micro-level' should be taken to refer to truths about the intrinsic protophenomenal nature of the component parts and their structural relations, and 'macro-level' should be taken to refer to truths about the phenomenal nature of the system as a whole:

Micro-level/Macro-level Gap: For any group of parts, $P_1, P_2 \dots P_n$ which instantiate protophenomenal properties, $Q_1, Q_2 \dots Q_n$ and stand in structural relation R , the following scenario is conceivable: there are $P_1, P_2 \dots P_n$ instantiating $Q_1, Q_2 \dots Q_n$ and standing in R , but where nothing of which $P_1, P_2 \dots P_n$ are proper parts of is conscious.

This gap can be defended in an analogous way to our Russellian panpsychist's Micro-level/Macro-level gap and again carries an equal epistemic threat.

We can use either the Protophenomenal/Phenomenal gap or the Micro-level/Macro-level gap to produce a Conceivability Argument against Russellian panprotopsyism; we can conceive of P^* – as defined in terms of physical fundamental parts with protophenomenal properties – but where there is no phenomenal experience simpliciter, or where there is no macro-experience. I shall set out the argument below, explain how either one of our epistemic gaps can be used to support P.1, and then defend the argument in its conclusion. In this instance, and similar to case with Russellian panpsychism, since the Russellian panprotopsyist has redefined P in terms of properties relevant to instantiating Q , P now has properties relevant to Q (and its lack of primary and secondary intensions) characterising it. As with Russellian panpsychism, the only option to resist this would be to propose that the intrinsic nature of P does not coincide with properties relevant to Q , but this would again remove both the positive component of her thesis: providing P with an intrinsic nature which resolves the issues associated with causal structuralism, and which also explains why it is that we are conscious.

Just as we did in our arguments for Russellian panpsychism, we can use broadly physical truths to refer to narrowly physical truths *plus* quiddities, but where quiddities are now defined in terms of protophenomenal properties. Accordingly, ' P^* ' should now be taken to represent the complete broadly physical truths about the universe (as defined in terms of protophenomenal properties), and ' Q ,' should be taken again to represent macro phenomenal truths about the universe.

- P.1 It is conceivable that $P^* \& \sim Q$.
- P.2 If it is conceivable that $P^* \& \sim Q$, it is primarily possible that $P^* \& \sim Q$.

P.3	If it is primarily possible that $P^* \& \sim Q$, then $P^* \& \sim Q$ is secondarily possible.
P.4	If it is secondarily possible that $P^* \& \sim Q$, it is metaphysically possible that $P^* \& \sim Q$, and so Russellian panprotopsyism is false.
C.	Russellian panprotopsyism is false.

Both the Protophenomenal/Phenomenal gap or the Micro-level/Macro-level can be used to defend the claim that $P^* \& \sim Q$ is conceivable. In the case of the former gap, we can conceive of a broadly physical isomorph – as described in terms of its narrowly physical properties plus protophenomenal properties – but where that isomorph has no phenomenal experience. In the case of the latter gap, we can conceive of a broadly physical isomorph – as described in terms of the intrinsic protophenomenal nature of its component parts, alongside its structural relations – but where the isomorph has no macro-experience. Since both gaps have already been argued to satisfy the notion of what is ideally, negatively, primarily conceivable to the degree required by Chalmers, I shall refrain from restating these arguments.

P.1 reemphasises the existence of an epistemic gap between P^* and Q . If it were the case that P^* epistemically entailed Q , if P^* , then Q , necessarily. Through qualifying the conceivability of $P^* \& \sim Q$ as ideal negative conceivability, we establish the claim that there will never be entailment between P^* and Q such that P^* a priori explains Q , or that Q is deducible from or fully explainable by P^* . P.2 then acknowledges the primary possibility of $P^* \& \sim Q$. The apparent lack of distinct primary and secondary intensions within the statement has then supported the claim of P.3: that the primary possibility of $P^* \& \sim Q$ entails the secondary possibility of $P^* \& \sim Q$ and so there is a possible world in which the secondary intension of $P^* \& \sim Q$ is true. Finally, since Russellian panprotopsyism requires that P^* constitutively grounds Q , it requires that if P^* obtains, Q necessarily also obtains in any possible world. However, per P.3, $P^* \& \sim Q$ is

secondarily possible. Since this acknowledges the possibility of a world in which P^* does not necessitate Q , P^* must not then necessitate Q , and so Russellian panprotopsyism must be false.

Like before, in the absence of a successful positive proposal to avoid the aforementioned epistemic gaps, the only real available objection to the Russellian panprotopsyist is to appeal to ignorance. For example, she might claim that we don't have a positive understanding of what protophenomenal properties are, nor how they interact, and so we cannot be fairly tasked to determine what is a priori deducible from them. However, there is little reason to grant this response any greater dialectical force than we did to her opponents. In light of this prior analysis then, we should say that if the conclusion holds in Conceivability Arguments against physicalism or Russellian panpsychism, it also holds here against our Russellian panprotopsyist; granting P.1-P.4, Russellian panprotopsyism is false.

Following Richard Brown (2010), we can also put forward similar epistemic gaps against dualism, which can in turn be formulated into Conceivability Arguments. First, let us consider property dualism. Recall, this thesis endorses ontological monism and property pluralism, whereby there is only one substance – the physical – but two fundamentally distinct types of property: physical properties and mental properties (of which phenomenal experience is one). I include property dualism alongside traditional dualism here simply to ensure that the general spectrum of philosophy of mind hypotheses are accounted for in my analyses; it will help finalise the charge that Conceivability Arguments can be applied to *any* philosophy of mind hypotheses.

Now, like each of our previous theses, by positing a relation between potentially qualitatively different things, Brown proposes that the property dualist also makes himself vulnerable to a new kind of entity: an ‘np-zombie.’ To conceive of this kind of entity, we should attempt to imagine a creature that is non-physically identical to ourselves in every respect, but which lacks either the entirety or some aspect of consciousness, for example seeing the colour red (Brown, 2010: 50). Via the same charge provided to our previous theses, there is no obvious contradiction that emerges in this situation obtaining. Brown proposes that we can furnish our conception of this kind of entity with the following potential epistemic gap:

Nonphysical/Phenomenal Gap: there is no a priori entailment from nonphysical truths to phenomenal truths.

(ibid).

To understand this gap, one must first concede the possibility of nonphysical truths existing. We must then imagine the totality of nonphysical properties that could exist being instantiated within some system (in whatever way they would be instantiated, were their reality assured), but without any phenomenal properties. More formally, we can redefine the gap as follows:

For any group of nonphysical properties $NP_1, NP_2...NP_n$ standing in some nonphysical relation NR, the following scenario is conceivable: there are $NP_1, NP_2...NP_n$ standing in NR, but it’s not the case that $NP_1, NP_2...NP_n$ standing in NR is phenomenally conscious.

Now, one might have immediate concern about the conceivability of such an entity. As noted, property dualism proposes that phenomenal properties are nonphysical properties, and so – granting property dualism – one cannot have the former class of property without also having the latter; they are allegedly one and the same. However, Brown proposes that there is a clear

sense in which the reasoning behind this epistemic gap mirrors those proposed against physicalism. In fact, np-zombies very closely resemble p-zombies: the traditional p-zombie is just a creature that has everything I do physically but lacks phenomenal properties, and an np-zombie is just a creature that has everything I do non-physically but lacks phenomenal properties (Brown, 2009: 50). In this manner, while it may appear that Brown has simply assumed that phenomenal properties are not nonphysical properties – which seems unfair to the property dualist – this precise form of complaint can also be made by physicalists with regards to the assumption against them that phenomenal properties are not physical (ibid: 51). While this line of objection will be returned to later, let us for the moment then grant this epistemic gap, and input it into a parallel Conceivability Argument against property dualism.

‘NP’ should be taken to represent the complete nonphysical truths about the universe (including all nonphysical properties) and ‘Q,’ should be taken to represent phenomenal experience. Again, we can say that the primary and secondary intensions of NP and Q should coincide; NP has been defined in terms of properties relevant to instantiating Q, and so NP now has properties relevant to Q (and its lack of primary and secondary intensions) characterising it.

- | | |
|-------|---|
| P.1 | It is conceivable that NP&~Q. |
| P.2 | If it is conceivable that NP&~Q, it is primarily possible that NP&~Q. |
| P.3 | If it is primarily possible that NP&~Q, then NP&~Q is secondarily possible. |
| P.4 | If it is secondarily possible that NP&~Q, it is metaphysically possible that NP&~Q, and so property dualism is false. |
| <hr/> | |
| C. | Property dualism is false. |

(Brown, 2010: 50).

P.1 reemphasises the existence of an epistemic gap between NP and Q. If it were the case that NP epistemically entailed Q, one could deduce Q from their knowledge of NP. Through qualifying the conceivability of NP&~Q as ideal negative conceivability, we establish the claim that there will never be sufficient entailment between NP and Q such that NP a priori explains Q, or that Q is deducible from or fully explainable by NP. P.2 then acknowledges the primary possibility of NP&~Q. The apparent lack of distinct primary and secondary intensions within the statement has then supported the claim of P.3: that the primary possibility of NP&~Q entails the secondary possibility of NP&~Q and so there is a possible world in which the secondary intension of NP&~Q is true. Finally, since property dualism requires that NP constitutively grounds Q, it requires that if NP obtains, Q necessarily also obtains in any possible world. However, per P.3, NP&~Q is secondarily possible. Since this acknowledges the possibility of a world in which NP does not necessitate Q, NP must not then necessitate Q, and so property dualism must be false.

Let us first acknowledge that this argument parallels all of the Conceivability Arguments I have put forth so far, and so granting their validity, alongside the truth of the premises here, the conclusion against property dualism holds. Nonetheless, the property dualist might still attempt to respond in a number of ways. As noted, the property dualist might take issue with the idea that the totality of nonphysical properties can exclude certain or all phenomenal properties. First and foremost, the property dualist proposes a dualistic account of properties so as to account for phenomenal properties, and so any account of NP should then account for *all* phenomenal properties. Crucially, however, one will only find this response convincing if they first assume that phenomenal properties are nonphysical and this has not yet been established (Brown, 2010: 63).⁵⁸ Indeed, the dualist can only provide such a response *if* they have already

⁵⁸ This will be returned to in Section 4.4.2.

established their own conclusion – which is the very matter at hand – and at which point, the task of Conceivability Arguments would be impotent. Accordingly, the critic can simply charge them with assuming the very thing that they are trying to prove.

To avoid begging the question, Brown proposes that the property dualist might instead challenge the soundness of the claim that $NP \& \sim Q$ (ibid: 62). To do this, they might propose that although nonphysical properties are necessary for phenomenal experience, they might not be sufficient in-and-of-themselves. For example, he might propose that such nonphysical properties remain inert until they are instantiated within some certain physical process, i.e., that they require those entities of which they are a property of be physically realised and organised in an appropriate fashion. However, responses of this form must also falter. Our NP-zombie is (only) nonphysically identical to ourselves, in such a manner that the argument permits that there can be certain physical properties or processes present so as to realise those nonphysical properties, if in fact they were necessary to do so: it is the nonphysical properties that we are keeping fixed, not the lack of physicality. In this manner, any nonphysical phenomenal properties we have, our NP-zombie must also have: if we do not have inert properties, then neither does our NP-zombie. Accordingly, any world which is proposed to contain inert nonphysical properties is not an NP-zombie world. Since there cannot then be inert nonphysical properties in our NP-zombie world, the claim that $NP \& \sim Q$ is unsound fails. With these potential responses considered, and via the same reasoning of the previous subsections, the conclusion of property dualism's falsity must then follow.

Let us now turn to substance dualism: the thesis that mind is a distinct substance to the physical. In many ways, the formulation of a Conceivability Argument against substance dualism (henceforth, dualism), is easier than the case for property dualism. In the same way that the

metaphysical possibility of P and not Q would negate the truth of physicalism, so too can P *and* Q negate the truth of dualism: if we can conceive of a scenario in which the phenomenal logically supervenes on the physical, we can establish the primary and so secondary possibility of such a world, and so show the modal thesis of dualism to be false.

With this in mind, let us consider a new type of a p-zombie, again introduced by Richard Brown (2010). I shall call it a ‘~zombie.’ Now, a ~zombie is a creature which is (only) physically identical to me, viz. he is molecule for molecule identical to myself in all physical capacities, is situated within an identical physical environment and social scenario, and crucially, has nothing more (nonphysically) than my physical constitution and situation. Now, unlike our traditional p-zombie, our ~zombie *has* phenomenal experience: he feels pressure under his fingertips as he types, enjoys the fragrant satisfaction of his sipping coffee, and is understanding each and every word he writes in his ~zombie thesis. For our ~zombie there *is* something it is like to be him. Crucially, however, in this case, the phenomenal experience is entirely grounded in the physical, such that it is nothing over and over it.

Let us formalise a final Conceivability Argument. ‘P’ should be taken now to represent the complete narrowly physical truths about the universe, and ‘Q,’ should be taken to represent phenomenal truths about the universe. Since this Conceivability Argument entails both P and Q obtaining, ‘ $\supset$ ’ should be taken here to refer to grounding by analysis, such that if $P \supset Q$, then P grounds Q in such a manner that Q is nothing more than P. It should be clear again here that the primary and secondary intensions of sentences involving these variables coincide. Q has been argued to possess primary and secondary intensions that coincide, and P is defined here by the physicalist who endorses causal structuralism. Granting this, all that a property is, is

what it does, and so the primary and secondary intensions of P are the same. Accordingly, we can formulate the following argument:

P.1	It is conceivable that $P \supset Q$.
P.2	If it is conceivable that $P \supset Q$, it is primarily possible that $P \supset Q$.
P.3	If it is primarily possible that $P \supset Q$, then $P \supset Q$ is secondarily possible.
P.4	If it is secondarily possible that $P \supset Q$, it is metaphysically possible that $P \supset Q$, and so traditional dualism is false.
C.	Traditional dualism is false.

As before, let us again acknowledge that this argument parallels all of the Conceivability Arguments I have put forth so far, and so granting their validity, alongside the truth of the premises here, the conclusion holds. P.1 emphasises that one can conceive of P epistemically entailing Q. Through qualifying the conceivability of $P \supset Q$ as ideal negative conceivability, we establish the claim that we could never rule out a world in which P grounds Q. P.2 then acknowledges the primary possibility of $P \supset Q$. The apparent lack of distinct primary and secondary intensions within Q has then supported the claim of P.3: that the primary possibility of $P \supset Q$ entails the secondary possibility of $P \supset Q$. Since the sentence $P \supset Q$ lacks distinct primary and secondary intensions, there is a possible world in which the secondary intension of $P \supset Q$ is true. Finally, since dualism requires that P does not constitutively ground Q, it requires that if only P obtains in some possible world, Q does not. However, per P.3, $P \supset Q$ is secondarily possible. Since this acknowledges the possibility of a world in which P does necessitate Q, some distinct nonphysical substance over and above P must not then necessitate Q, and so dualism must be false.

Brown acknowledges that the dualist might take issue with the argument by responding that it merely amounts to a brute assertion about physicalism's truth. By this, one might propose that the argument is really only proposing the following:

- (1) It is conceivable that there could be a physically identical world to ours in which physicalism is true.
- (2) Since physicalism makes a modal claim, it necessitates that what is true of any world physically identical to our own, it is true of our world.
- (C) Physicalism is true of our world.

(Brown, 2010: 64).

Brown provides a number of possible rebuttals to this form of response. First, since our $\sim$ zombie Conceivability Argument mirrors our Complex Conceivability Argument against physicalism, the latter argument must then also amount to a brute assertion about physicalism's falsity. However, since the proponent of Conceivability Arguments has already dismissed such reasoning, it cannot hold here in their defence (ibid). Second, although the claim that ' $P \supset Q$ is conceivable' matches the physicalist's thesis, it is not making a 'meta-modal' claim about the entire space of possibilities and so is not making a claim about the modal thesis of physicalism (ibid: 65).⁵⁹ Instead, it is simply conceiving of one possible world in which ' $P \supset Q$ ' obtains. This is not to assume that physicalism as a modal thesis is true, but rather that physicalism is metaphysically possible and so that dualism, as a modal thesis, is false: the falsity of dualism here has no bearing on the truth of physicalism within the purview of this argument. In this manner, one is not using a brute assertion about physicalism to prove physicalism but rather using the metaphysical possibility of physicalism to disprove dualism. With this in mind, and

⁵⁹ 'Meta-modal' refers to the higher-order reflection upon modal claims, such as the modality of modal properties.

via the reasoning of our previous subsections, the conclusion of dualism's falsity must also then follow.

As we have seen, each of the available dualist responses centre upon the notion that the above Conceivability Arguments beg the question against dualism. However, as briefly touched upon – and as we shall also come to see in greater detail in Section 4.4.2 – one cannot make this charge without acknowledging that each of the previous Conceivability Arguments also beg the question against the theory they are formulated against.⁶⁰ In this manner, we should say that if the Conceivability Arguments against physicalism, Russellian panpsychism, and Russellian panprotopsyism hold, then so too do they hold for dualism.

As documented in this thesis, many authors have posited Conceivability Arguments against a wide range of philosophy of mind hypotheses. However, none have yet entertained the potential consequences of this breadth of application, especially with regards to what this means for all of those philosophy of mind hypotheses, or for our use of the Conceivability Argument itself. Granting the equal dialectical force of my proposed arguments, and again assuming that the basic argument is valid, we can harness this breadth of application to consider the following three potential consequences. These are not logically exhaustive, but simply representative of three clear possibilities we can draw from my analysis.

- (i) All of the arguments' premises are true in each application, and our entire spectrum of hypotheses are incorrect. Or,
- (ii) At least one of their premises is false in each application, and so they cannot provide us with a true conclusion about any philosophy of mind hypothesis. Or,

⁶⁰ I refrain from delving into such reasoning yet, only to strengthen the overall claim of this chapter: there are a few more variables and conceivable situations that must first be set up before we can truly illuminate the error that is occurring in our application of Conceivability Arguments.

- (iii) An adequate justification for the truth of at least one of their premises in each application is currently unattainable, and so they are currently of no use within the Mind-Body debate.

I am not claiming here that one should extend these conclusions to Conceivability Arguments, *simpliciter*. Indeed, simply because one can produce, for example, silly *reductio ad absurdum* arguments, they should not extend that *all* reductio ad absurdum arguments are silly. Rather, I am claiming that dependent on one's antecedent assumptions, one may find each of my arguments to be as justified as the last. Assuming that the basic argument structure is valid, the proponent of any of these stated Conceivability Arguments is presented with these three potential consequences. Crucially, those who would claim one or the other argument to be unsound should only help to reemphasise those antecedent assumptions at work.

4.4 ~ The Issue with Conceivability Arguments

In this section, I will construct three 'Counterfactual Conceivability Arguments,' (or 'CCAs') which use Conceivability Arguments to negate a metaphysical *truth* about the world from which they are posited. For the sake of simplicity, these shall be worlds in which physicalism is true.⁶¹ The purpose of this analysis will be to illustrate that Conceivability Arguments can be validly used to establish a contradictory state of affairs and could be telling us things that contradict the truth about our world. In this manner, I shall propose that there must currently be an issue with at least one premise within each of my CCAs. This, alongside the fact that it appears unlikely that our entire spectrum of philosophy of mind hypotheses should turn out to

⁶¹ It should be noted that I shall use physicalism in this section as a placeholder: the counterfactual states of affairs I address centre around physicalism for the sake of simplicity, but further CCAs could just as easily be applied to worlds about which forms of dualism or Russellian monism is true.

be incorrect will lead me to the conclusion that consequence (i) should not be drawn from my analysis: one should not maintain that all of the arguments' premises are true in each application, and that our entire spectrum of hypotheses are incorrect. In this manner, it is not the case that Conceivability Arguments against physicalism, Russellian monism, or dualism proves those theories to be false.

Continuing under the assumption that each of my arguments are equivalently strong, both remaining consequences necessitate that a fault lies within at least one premise of each argument. In Section 4.4.2, I shall appeal to my CCAs to locate this fault primarily in P.1, specifically with regards to the claim that any of the stated scenarios are currently ideally conceivable. I shall propose that the required ideal perspective is currently unattainable and that appeals to it simply amount to brute assertions that beg the question against the theory one intends to attack. I will then locate a further fault in the claim that the primary and secondary intensions of Q coincide for the physicalist. Finally, I will harness these sub-conclusions to propose that consequence (iii) is the most likely. I will, therefore, recommend that Conceivability Arguments are relinquished from the Russellian panpsychist's repertoire. Doing so removes one the most popularly appealed to arguments for the theory, but, indeed, such a move is pertinent if she wishes to make a compelling case for her theory's superiority over physicalism.

4.4.1 Counterfactual Conceivability Arguments

My motivation for this subsection stems from Katalin Balog's thesis within 'Conceivability, Possibility, and the Mind-Body Problem' (Balog, 1999). Here, Balog suggests that if p-zombies – or that $P \& \sim Q$ – are genuinely metaphysically possible, then such p-zombies – or entities from within a world in which $P \& \sim Q$ obtains – could produce an analogous Conceivability Argument

to the ones we have used here. To begin, she asks us to suppose that the original 2-Dimensional Conceivability Argument is sound. The conclusion that physical facts do not necessitate phenomenal facts, would then be true (at least in that world); it is a minimal physical duplicate world in which no phenomenal, or indeed other nonphysical facts obtain (Balog, 1999: 512). In this manner, we can say that this is a world in which physicalism is necessarily true (ibid: 513). Since this a world in which physicalism is necessarily true, our p-zombie is an entity about which physicalism is also necessarily true.

Harnessing this conclusion, Balog asks us to conceive of our p-zombie also proposing a Conceivability Argument. After all, he is a minimal physical duplicate, and so it is conceivable that he is putting forward an identical argument to that which we did. For example, our p-zombie could also propose that it is conceivable that $P \& \sim Q$, where the concepts of our p-zombies argument that correspond to Q in Chalmers' argument simply now refer to some (physical) state (ibid). Continuing to assume that the basic argument is sound, our p-zombie would establish an analogous Conceivability Argument which negates the modal truth of physicalism in his world. Crucially, and as noted, in our initial Conceivability Argument against physicalism, the conceived world in which that $P \& \sim Q$ obtains is a world in which physicalism is necessarily true: our p-zombie world is a minimal physical duplicate in which P is the sole ingredient and so is a world which is, by definition, physicalist (ibid: 503-504). In this manner, Balog suggests that we can have a genuinely physicalist world from within which Conceivability Arguments show physicalism to be false.⁶²

⁶² One might be immediately puzzled of what 'Q' could refer to in a zombie world. This worry shall be described in more detail shortly with reference to Chalmers' own criticism of Balog.

By Balog's logic, there must be something inherently wrong with Conceivability Arguments: physicalism is necessarily true in her p-zombie's world, and yet we have a Conceivability Argument from within that world which allegedly establishes the contrary. Furthermore, since this second Conceivability Argument mirrors the initial Conceivability Argument against physicalism, its conclusion must be treated with the same validity. What we are left with, is a clear contradiction: conceivability – and its link to possibility – has established a world in which physicalism is true, but where those same links between conceivability and possibility about those same variables from within that world have established that physicalism is false (ibid: 510-514). By virtue of this contradiction, and due to the arguments being identical, Balog proposes that there must then be a false premise within *both* arguments.

Granting Balog's analysis, consequence (i) of my analysis would be discounted. Furthermore, by virtue of establishing false premises in both arguments, she would likely adopt consequence (ii) as the most appropriate. Now, since Balog's analysis focusses on the Simple Conceivability Argument, she locates this false premise as P.2: the inference from conceivability to possibility. Since I am working with the Complex Conceivability Argument, it may be useful to the reader to acknowledge that Balog would probably locate the analogous fault as somewhere within either P.2 and P.3, or within both. My reasoning for this is solely due to P.2 and P.3 of the Complex Conceivability Argument being those that link conceivability to possibility, whereas in the Simple Conceivability Argument, only P.2 deals with this link.

Chalmers has since responded to Balog that it need not be the case that just because the premises of the original Conceivability Argument against physicalism are true, that the premises of our p-zombies Conceivability Argument are also true. He proposes that a tacit premise of the original argument is Q itself, or that, for example, 'someone is phenomenally

conscious' (Chalmers, 2010: 159). Indeed, such a premise is arguably needed to include Q as a variable at all. In the original 2-Dimensional Argument, we uphold the tacit premise ourselves; we are phenomenally conscious. Since the arguments are allegedly analogous, the p-zombies Conceivability Argument must then also contain the tacit premise 'someone is phenomenally conscious.' However, Chalmers proposes that this tacit premise is surely false in our p-zombie's Conceivability Argument: our p-zombie world is a world in which no one is phenomenally conscious, and so it is not a world in which the tacit premise that someone is phenomenally conscious can be upheld. Per Chalmers, the p-zombies' argument therefore has a false premise where the original argument has a true premise, and so the conclusion of the p-zombies argument does not undermine the original argument (ibid: 160). Arguably then, there is no real contradiction established via Balog's example, for the p-zombie's argument does not establish a genuine metaphysical truth.

I am sympathetic to Chalmers' objection. There appears to be a clear sense in which our p-zombie cannot validly make any claims about Q or $\sim Q$, since they live in a world in which Q simply does not obtain at all. Firstly, Balog's p-zombie cannot meaningfully claim to have conceived of anything, since they are a p-zombie: they have no Q about which they can conceive. While Balog does qualify that Q in our p-zombie's argument should refer to some (physical) state, it is intuitive that without our p-zombie upholding Q tacitly, there is nothing about which to conceive of in the first place (Balog, 1999: 513). Secondly, one needs to have Q, and a concept of that Q in order to negate the truth of physicalism in that world: one cannot make conceivability claims about Q without having the concept of Q to input into claims regarding it. Therefore, it must be false that a p-zombie can conceive of a p-zombie. In this light, the contradiction displayed in Balog's argument is not indicative of a false premise within Conceivability Arguments simpliciter, but rather only a false premise within her own argument.

Although I am sympathetic to Chalmers, I believe that Balog's criticism can be reformulated in such a way that it becomes successful. For the remainder of this subsection, I will reconstruct three Balog-esq scenarios which endeavour to bypass Chalmers' concerns. Continuing under the assumption that the argument structure is valid – and the typical defence of its related premises sufficient – I shall produce a series of genuine contradictions from Conceivability Arguments. I will then harness these contradictions to illustrate that it cannot be the case that all of the Conceivability Arguments' premises are true in each application, and our entire spectrum of hypotheses are incorrect. While the first argument will be sufficient to establish Balog's intended conclusion regarding flawed or false premises – and so to dispel consequence (i) – the latter two will be used in the next subsection to decipher which of the remaining consequences – (ii) or (iii) – is the most appropriate, alongside where within the argument this flawed or false premise resides. Contra to Balog, I shall not claim P.2 to be false. Instead, I shall conclude that P.1 is *flawed*, specifically with regards to the claim that any of the stated scenarios are ideally conceivable. In this manner, I shall ultimately conclude consequence (iii), rather than consequence (ii), to be most appropriate.

To begin my first Counterfactual Conceivability Argument, we must first set up the notion of 'partial zombies' (Chalmers, 1996: 133; 2010: 107). Now, partial zombies are an amalgamation between the notion of a p-zombie, and a ~zombie. Recall, the former refers to an entity which is physically and behaviourally identical to us, but that lacks consciousness entirely; the latter refers to a mere physical duplicate to ourselves – viz. has nothing more than the physical properties we have – and which has phenomenal consciousness. Where the former feels no pressure under his fingertips while he types his zombie thesis, the latter feels pressure; where

the former enjoys no fragrant satisfaction from his sipping coffee, the latter does; and, where the former has nothing it is like to be him, the latter has something it is like to be him.

Now, like both our $\sim$ zombies and p-zombies, our partial zombie is physically and behaviourally identical to ourselves: he is a minimal physical duplicate. Like our $\sim$ zombie, there is also something it is like to be him. However, unlike our $\sim$ zombie, Q is not instantiated to the level which it is instantiated within us. In particular, we can say that he lacks *some* particular aspect of consciousness or Q (that we ourselves have) but enjoys all the rest. Perhaps, what he is missing is the phenomenal experience of colour, or of taste, or of smell. Now, this lack of particular phenomenal experience should not be treated as analogous to typical cases of monochromacy, aguesia, or hyposmia, respectively. In these instances, there is a physical ground for the lack of ability to experience colour, taste, and smell. Our partial zombie, however, is molecule for molecule identical to those who *can* experience, in all physical capacities and is also situated within an identical physical environment and social scenario. In this manner, our partial zombie is a minimal physical duplicate (such that P is the sole ingredient), but where there is less Q entailed than that in a $\sim$ zombie.

Let us entertain the notion of what this scenario might entail. My partial zombie twin would be behaviourally and functionally identical to me: he would receive the same sort of information, process it in an appropriate and cognitively indistinguishable way to myself, and produce indistinguishable resultant behaviour. Like myself, he is tapping away on his computer, sipping coffee, and introducing philosophical notions. Like myself, there *is* phenomenal experience of pressure under my p-zombie twin's fingertips and an understanding of the notions he is introducing to his thesis. However, let us say that, unlike me, he enjoys no fragrant satisfaction from his sipping of the coffee since he has no phenomenal experience of taste, nor even any

comprehension of what it would be like to taste. In this manner, our partial zombie is entirely unaware that he is a partial zombie.

Like our previous zombie notions, there is little reason to believe that partial zombies exist in the actual world. I know that I can taste, and I can extend that – at least under this planet’s laws of nature – entities of a relevantly similar physical constitution and behaviour to myself also enjoy this experience. However, if p-zombies are conceivable, then partial zombies certainly are too. Unlike p-zombies, the notion of partial zombies does not require an absence of Q, but rather only that Q is not instantiated to the degree that we ourselves enjoy. Indeed, depending on one’s antecedent assumptions, it might appear more coherent that one can conceive of all of the P without *all* the Q, than it is for one to conceive of all of the P without *any* of the Q. Where the former conception can be inputted into a Conceivability Argument to suggest that Q does not supervene on P at all, the latter can be inputted to suggest that some or most aspects of Q might supervene on P, but that some aspects may not.

To say that partial zombies are conceivable here is to again say that it is ideally negatively, primarily conceivable that P *grounds* Q^{-1} . As before, for this conceivability to be satisfied, it must be the case that even from a perspective of ideal reasoning, no amount of a priori reasoning could deduce that the primary intension of P grounding Q^{-1} is incoherent. The case for this is easier than it was for our initial p-zombie: it must just be that some aspect of the epistemic gap between P and Q remains even with idealised knowledge of P and Q. While a case could be made for this via similar reasoning to our considerations about the Hard Problem, we need not flesh them out here. Instead, it will suffice to say simply that if $P \& \sim Q$ is conceivable, then $P \supset Q^{-1}$ must then also be; if p-zombies are conceivable, the conceivability of partial zombies is also entailed.

‘P’ should be taken to represent the complete narrowly physical truths about the universe, and ‘Q⁻¹’ should now be taken to represent *nearly* all of the phenomenal truths about the universe, i.e., where some particular truth is missing. As before, ‘ $\supset$ ’ should be taken here to refer to grounding by analysis, such that if $P \supset Q^{-1}$, then P grounds Q⁻¹ in such a manner that Q⁻¹ is nothing more than its grounds.

With this set up, we can produce a partial zombie Conceivability Argument:

- | | |
|-------|--|
| P.1 | It is conceivable that $P \supset Q^{-1}$. |
| P.2 | If it is conceivable that $P \supset Q^{-1}$, it is primarily possible that $P \supset Q^{-1}$. |
| P.3 | If it is primarily possible that $P \supset Q^{-1}$, then $P \supset Q^{-1}$ is secondarily possible. |
| P.4 | If it is secondarily possible that $P \supset Q^{-1}$, it is metaphysically possible that $P \supset Q^{-1}$, and so physicalism is false. |
| <hr/> | |
| C. | Physicalism is false. |

While this argument could be used to negate the modal truth of physicalism, as we have already levelled such arguments, we shall not use it for this purpose here. Instead, what is more important is the metaphysical possibility of a world in which $P \supset Q^{-1}$. In a similar way that Balog’s zombie world was a world about which physicalism was true (in that P was the sole ingredient of the world), our partial zombie world is also a world about which physicalism is true: it is a minimal physical duplicate world in which all of the Q that exists within that world is grounded by P. For this world, it is irrelevant that there could – in another possible world – be more Q than that which that world instantiates. For this world, if P, then Q⁽⁻¹⁾, necessarily. Crucially, this is not to say that the modal claims of physicalism are true; whatever aspect of Q our partial zombie lacks is not entailed by its possession of P, and so the modal claims of physicalism would be false. However, this is a possible world *about which* physicalism is true,

in the sense that it is a minimal physical duplicate in which P is the sole ingredient, and which transparently accounts for all of the Q in that world.

With this physicalist world set up as a genuine metaphysical possibility, we can now introduce the first of my three Counterfactual Conceivability Arguments. As noted, the main purpose of these arguments is to show how Conceivability Arguments can be used to negate a metaphysical truth about the world from which they are posited. However, they will also come to be instrumental in my final analysis regarding where the issue in Conceivability Arguments lies. After setting up all three Counterfactual Conceivability Arguments, I will then explain how Conceivability Arguments fail to tell us anything about the world in which we live, and so that we should not accept the first potential consequence of my analyses: that all of the arguments' premises are true in each application, and our entire spectrum of hypotheses are incorrect.

Since our partial zombie only lacks some aspect of Q, it is possible that he too can conceive of scenarios or states of affairs. Perhaps, in this partial zombie world, there is a partial zombie philosopher putting forward Conceivability Arguments with the hope to learn something about the state of affairs regarding consciousness on his world. Now, unlike in Balog's example, our partial zombie world *is* a world in which people are phenomenally conscious, and so is a world in which the tacit premise 'someone is phenomenally conscious' can be upheld. Accordingly, we can bypass Chalmers' concerns about false tacit premises. For example, if our partial zombie lacks the phenomenal experience of taste but has all of the other Q (and all of the P), he can conceive of states of affairs in which P obtains, but where Q varies to some degree with regards to his own. In this manner, he could conceive of a partial zombie (relative to himself) – one which lacks not just the ability to taste, but also to, for example, smell. For the sake of

simplicity, and to avoid any added confusion, let us call this state of affairs that $P \supset Q^{-2}$. Crucially, however, since our partial zombie does not know he is a partial zombie, the state of affairs depicted as $P \supset Q^{-2}$ would be depicted by himself as $P \supset Q^{-1}$, and so his Conceivability Argument would mirror our own in both its formulation and variables; *our partial zombie is simply conceiving of a partial zombie, relative to himself*. Now, if our partial zombie can conceive of a world in which $P \supset Q^{-2}$, he can produce a Conceivability Argument about $P \supset Q^{-2}$ that negates the truth of physicalism. With this in mind, we can formulate our first Counterfactual Conceivability Argument:

Counterfactual Conceivability Argument 1 (CCA1):

P.1	It is conceivable that $P \supset Q^{-1}$.
P.2	If it is conceivable that $P \supset Q^{-1}$, it is primarily possible that $P \supset Q^{-1}$.
P.3	If it is primarily possible that $P \supset Q^{-1}$, then $P \supset Q^{-1}$ is secondarily possible.
P.4	If it is secondarily possible that $P \supset Q^{-1}$, it is metaphysically possible that $P \supset Q^{-1}$, and so it is possible that there is a world about which physicalism is true.
C.1	There is a possible physicalist world, in which $P \supset Q^{-1}$.
P.5	If $P \supset Q^{-1}$ is metaphysically possible, it is possible that there is a partial zombie ($P \supset Q^{-1}$), who can conceive of a partial zombie ($P \supset Q^{-2}$).
P.6	If a partial zombie ($P \supset Q^{-1}$) can conceive of that $P \supset Q^{-2}$, he can produce a Conceivability Argument about $P \supset Q^{-2}$, which shows physicalism to be false.
C.2	There is a possible physicalist world (in which $P \supset Q^{-1}$) in which Conceivability Arguments show physicalism to be false.

Following CCA1, we have a possible world about which physicalism is true, but where a further Conceivability Argument from within that world shows physicalism to be false. In this manner, the conclusion of the partial zombie's Conceivability Argument contradicts the

conclusion of our initial partial zombie Conceivability Argument. Accordingly, there must be a flawed premise within at least one of the arguments. However, since our partial zombie's Conceivability Argument appears to mirror the premises of our initial partial zombie Conceivability Argument, there might be a flawed premise within *both arguments*, such that we can establish Balog's intended conclusion.⁶³

Rather than delving into further analyses just yet, let us move onto our second Counterfactual Conceivability Argument. Within this, our protagonist will again be a minimal physical duplicate – viz. molecule for molecule identical to myself in all physical respects – but who possesses all of the Q instantiated within our own phenomenal experience, with the addition of some extra qualitative capacity. We can call him a $\sim\text{zombie}^{+1}$ and represent this state of affairs as $P \supset Q^{+1}$ (where P grounds that Q^{+1}). Here, $P \supset Q^{+1}$ should be taken to refer to a world or entity which has all of the P and Q that a $\sim\text{zombie}$ has, but where Q is manifested in some further degree; since a $\sim\text{zombie}$ is qualitatively akin to ourselves, Q^{+1} must be some form of qualitative experience that we ourselves do not ourselves enjoy. Furthermore, and in a similar way to our initial $\sim\text{zombie}$ world, the $P \supset Q^{+1}$ world is a world in which physicalism is true: it is a minimal physical duplicate world in which all of the Q that exists within that world is grounded by P.

Now, it might appear difficult to conceive of a state of affairs in which an entity possesses aspects of qualitative experience over and above our own, since our worldview and the limits of our comprehension are constrained by our own subjective life. However, I believe there is a clear sense in which Q^{+1} remains negatively conceivable: we cannot rule it out. Furthermore,

⁶³ I say 'might' here, as one may take issue with the claim that the arguments mirror each other. If we separated the two arguments, they would – at face value – mirror each other, but the value of Q in $P \supset Q^{-1}$ and $P \supset Q^{-2}$ will still differ slightly in those possible worlds. Nonetheless, if we had no issue with Balog's zombie conceiving of a zombie (aside from those issues already mentioned), then we should have no issue with a partial zombie conceiving of a partial zombie.

we might even be able to make a positive conception if we consider animals in our world which have different qualitative experience to ourselves. For example, perhaps Q^{+1} refers to all of the Q we have in this world, with the additional ability to feel magnetic poles.⁶⁴ With this conception set up, we can say that it is conceivable that $P \supset Q^{+1}$, such that there is a possible physicalist world where $P \supset Q^{+1}$ is true.

Since our $\sim\text{zombie}^{+1}$ has more Q than that we ourselves enjoy, he can certainly conceive of scenarios or states of affairs in the same degree as ourselves. In his world, perhaps there is a philosopher putting forward Conceivability Arguments with the hope to learn about the state of affairs regarding consciousness in his world. Since our $\sim\text{zombie}^{+1}$ world is a world in which people are phenomenally conscious, it is a world in which the tacit premise that someone is phenomenally conscious can be upheld. Again then, we can bypass Chalmers' concerns about false tacit premises regarding the statement 'someone is phenomenally conscious' (such as those within Balog's argument).

Perhaps this $\sim\text{zombie}^{+1}$ philosopher also presents a Conceivability Argument where he conceives of a $\sim\text{zombie}^{+1}$ (relative to himself) – a minimal physical duplicate entity which possesses Q in all of the capacities instantiated within his own phenomenal experience, *plus* some extra qualitative capacity. So as to avoid confusion again, let us call this state of affairs that $P \supset Q^{+2}$. Again, however, since our $\sim\text{zombie}^{+1}$ does not know he is a $\sim\text{zombie}^{+1}$, the state of affairs depicted as that $P \supset Q^{+2}$ would be depicted by himself as that $P \supset Q^{+1}$, and so his Conceivability Argument would mirror our own in both its formulation and variables. Now, if our $\sim\text{zombie}^{+1}$ can conceive of a world in which $P \supset Q^{+2}$ obtains, he can produce a parallel

⁶⁴ Crucially, for the purposes of this discussion, this further qualitative capacity should be *only* a feeling; it should have no impact upon one's behaviour or else the world would no longer be a minimal physical duplicate.

Conceivability Argument about $P \supset Q^{+2}$ that negates the truth of physicalism in his world. Since it is genuinely the case in our $\sim\text{zombie}^{+1}$ world that Q^{+1} is grounded by analysis in P, it should be the case that if P, then Q^{+1} simpliciter: $P \supset Q^{+2}$ should be false. However, our $\sim\text{zombie}^{+1}$ has established a state of affairs in which this does not obtain, as $P \supset Q^{+2}$ has been established as a genuine metaphysical possibility.

With this in mind, we can formulate our second Counterfactual Conceivability Argument, where ' $\supset$ ' refers to grounding by analysis, such that if $P \supset Q^{+1}$, then P grounds Q^{+1} in such a manner that Q^{+1} is nothing more than P.

Counterfactual Conceivability Argument 2 (CCA2):

P.1	It is conceivable that $P \supset Q^{+1}$.
P.2	If it is conceivable that $P \supset Q^{+1}$, it is primarily possible that $P \supset Q^{+1}$
P.3	If it is primarily possible that $P \supset Q^{+1}$, then $P \supset Q^{+1}$ is secondarily possible.
P.4	If it is secondarily possible that $P \supset Q^{+1}$, it is metaphysically possible that $P \supset Q^{+1}$, and so it is possible that there is a world about which physicalism is true.
C.1	There is a possible physicalist world, in which $P \supset Q^{+1}$.
P.5	If $P \supset Q^{+1}$ is metaphysically possible, it is possible that there is some $\sim\text{zombie}^{+1}$ ($P \supset Q^{+1}$), who can conceive of a $\sim\text{zombie}^{+1}$ ($P \supset Q^{+2}$).
P.6	If some $\sim\text{zombie}^{+1}$ ($P \supset Q^{+1}$)-can conceive of that $P \supset Q^{+2}$, he can produce a Conceivability Argument about $P \supset Q^{+2}$ which shows physicalism to be false.
C.2	There is a possible physicalist world ($P \supset Q^{+1}$) in which Conceivability Arguments show physicalism to be false.

Like CCA1, we again reach a potentially contradictory state of affairs. We have established a possible world about which physicalism is true, but where a further Conceivability Argument

shows physicalism to be false in that world. As such, there must again be a flawed premise within at least one of the arguments. However, since the $\sim\text{zombie}^{+1}$ Conceivability Argument (P.5-C.2) appears to mirror our initial $\sim\text{zombie}^{+1}$ Conceivability Argument (P.1-C.1), there might also be a flawed premise within both arguments.

Let us now turn to our final Counterfactual Conceivability Argument. This will take a similar form to our previous two Counterfactual Conceivability Arguments, but the protagonist within it shall be an ‘ideal $\sim\text{zombie}$ ’. At face value, our ideal $\sim\text{zombie}$ is a $\sim\text{zombie}$ which possesses all of the P and all of the Q that we ourselves enjoy, with the addition of *all* of the additional possible Q. Now, an ideal $\sim\text{zombie}$ might – upon our possession of a complete knowledge – simply manifest as a phenomenal duplicate to ourselves (viz. it might simply be the case that level of phenomenal experience that we enjoy *is* the total amount of phenomenal experience possible to be grounded by our P). However, it might also not be the case: it might be the case that even from an ideal perspective, one can conceive of an entity which is physically identical to ourselves, but who possesses a vastly (or perhaps even infinitely) larger internal qualitative life, furnished with many further qualitative aspects. Nonetheless, a positive conception of such an entity is not necessary here, as we remain interested solely with negative conceivability. Since neither of these can be ruled out, we can conceive of either obtaining.

Let us entertain this notion in some more detail. First, we should again begin by conceiving of a normal $\sim\text{zombie}$: an entity which has all of the P and all of the Q which we ourselves enjoy. However, we should now add in extra aspects of Q to this entity in a similar fashion to the state of affairs depicted by $P \supset Q^{+1}$ in CCA2. Unlike the state of affairs depicted by $P \supset Q^{+1}$ in CCA2, however, we should not stop with one additional qualitative aspect. Instead, we should extrapolate the possibility of extra qualitative aspects to the n^{th} degree, such that whatever the

limit of Q is, that is what our ideal $\sim$ zombie has. If the state of affairs depicted by $P \supset Q^{+1}$ in CCA2 was difficult to comprehend, then our ideal $\sim$ zombie will certainly also be. However, one may satisfy their comprehension taking whatever notion of this extra qualitative aspect that we might be able to comprehend (for example, feeling magnetic poles), add it with any others that are available to the comprehension, and finally add in the abstract notion of further possible qualitative aspects that we will never ourselves come to know.

It may seem plausible to the reader that for any number of conscious states it will always be conceivable that there is a being with more, regardless of the fact that P is held fixed here. In some ways, the ' n^{th} degree' of relevance might be an infinity of infinities. However, since we continue to use a negative form of conceivability within Conceivability Arguments, infinite Q should not be an issue: we simply cannot rule out it obtaining. Let us say that to conceive of an ideal $\sim$ zombie is to conceive of $P \supset Q^{x^n}$. If we find this notion palatable, we can conceive of $P \supset Q^{x^n}$ and input it into a Conceivability Argument which establishes the metaphysical possibility of such an entity, and in doing so, establishes a final possible world about which physicalism is necessarily true.

Since our ideal $\sim$ zombie is furnished with all of the possible qualitative aspects from P , he can certainly conceive of scenarios or states of affairs in the same degree (or perhaps even to a more complex degree) as ourselves. In our ideal $\sim$ zombie world, perhaps there is an ideal $\sim$ zombie putting forward Conceivability Arguments with the hope to learn about the state of affairs regarding consciousness in his world. And, as before, since our ideal $\sim$ zombie world is a world in which people are maximally phenomenally conscious, we can again bypass Chalmers' concerns about false tacit premises regarding Q within Balog's argument. With this in mind, let us now consider our ideal $\sim$ zombie conceiving of $P \supset Q^{x^n}$, relative to himself. Now,

for our ideal $\sim$ zombie, $P \supset Q^{xn}$ refers to a world or entity which has all of the P and Q that our ideal $\sim$ zombie has, but where Q is manifested to a maximal amount. Our ideal $\sim$ zombie's Conceivability Argument should again mirror our own in both its formulation and variables: there is simply an ideal $\sim$ zombie conceiving of an ideal $\sim$ zombie relative to himself, whatever that might mean. For the readers benefit, however, we shall call this value Q^{xn+} .

Now, it might be immediately apparent to the reader that if our ideal $\sim$ zombie is to satisfy our strict definition of conceivability (and that conceivability is a reliable guide to possibility), he should only be able to conceive of an entity just like himself: he is an ideal $\sim$ zombie, and so has all of the Q possible to be manifested by his P.⁶⁵ Crucially however, this fact is unbeknownst to our ideal $\sim$ zombie. Indeed, we ourselves could be ideal $\sim$ zombies. In this manner, there appears to be a clear sense in which our ideal $\sim$ zombie's ignorance about the fact that he *is* an ideal $\sim$ zombie, alongside his ignorance of the truth of physicalism in his world, that permits his negative conception of a more qualitatively complex physical duplicate of himself. For the moment then, let us permit the possibility of our ideal $\sim$ zombie conceiving of such. Indeed, in this instance, it is only our knowledge of our ideal $\sim$ zombie's metaphysical situation that reveals the non-possibility of his conceived scenario.

With this concession made, let us allow our ideal $\sim$ zombie to input the conceivability of $P \supset Q^{xn+}$ into a Conceivability Argument which negates the modal truth of physicalism: if our ideal $\sim$ zombie can conceive of a world in which $P \supset Q^{xn+}$ obtains, he can produce a parallel Conceivability Argument about $P \supset Q^{xn+}$ that negates the truth of physicalism in his world. Since it is genuinely the case in our ideal $\sim$ zombie world that Q^{xn} logically supervenes on P, it should be the case that if P, then Q^{xn} necessarily: 'that $P \supset Q^{xn+}$ ' should be (and indeed is) false.

⁶⁵ This worry will be returned to in the next subsection.

However, our ideal $\sim$ zombie has allegedly established a state of affairs in which this does not obtain, as, by the logic of Conceivability Arguments, $P \supset Q^{xn+}$ has been established as a genuine metaphysical possibility from his own perspective. With this set up, we can now formulate our final Counterfactual Conceivability Argument. Again, ‘&’ should be taken to refer to grounding by analysis, such that if $P \supset Q^{xn}$, then P grounds Q^{xn} in such a manner that Q^{xn} is nothing more than P.

Counterfactual Conceivability Argument 3 (CCA3):

P.1	It is conceivable that $P \supset Q^{xn}$.
P.2	If it is conceivable that $P \supset Q^{xn}$, it is primarily possible that $P \supset Q^{xn}$.
P.3	If it is primarily possible that $P \supset Q^{xn}$, then $P \supset Q^{xn}$ is secondarily possible.
P.4	If it is secondarily possible that $P \supset Q^{xn}$, it is metaphysically possible that $P \supset Q^{xn}$, and so it is possible that there is a world about which physicalism is true.
C.1	There is a possible physicalist world in which $P \supset Q^{xn}$.
P.5	If $P \supset Q^{xn}$ is metaphysically possible, it is possible that there is some ideal $\sim$ zombie ($P \supset Q^{xn}$), who can conceive of an ideal $\sim$ zombie ($P \supset Q^{xn+}$).
P.6	If some ideal $\sim$ zombie ($P \supset Q^{xn}$) can conceive of that $P \supset Q^{xn+}$, he can produce a Conceivability Argument about $P \supset Q^{xn+}$ which shows physicalism to be false.
C.2	There is a possible physicalist world ($P \supset Q^{xn}$) in which Conceivability Arguments show physicalism to be false.

Like CCA1 and CCA2, we again reach a potentially contradictory state of affairs indicative of a shared flaw within the premises of both arguments. With these Counterfactual Conceivability Arguments now set up, we can now begin to make some remarks on the states of affairs they depict. For the moment, let us first turn to how they impact upon the three potential consequences of my analysis. Recall, these were that:

- (i) All of [my previous] arguments' premises are true in each application, and our entire spectrum of hypotheses are incorrect. Or,
- (ii) At least one of their premises is false in each application, and so they cannot provide us with a true conclusion about any philosophy of mind hypothesis. Or,
- (iii) An adequate justification for the truth of at least one of their premises in each application is currently unattainable, and so they are currently of no use within the Mind-Body debate.

Now, it appears unlikely from the offset that (i) is the correct consequence to accept. If physicalism, dualism, and everything in between them is incorrect, then we are left with little way to progress: we might simply have to conclude that Q is unexplainable. Now, while this may come to be the case, it is certainly not a conclusion that we should be drawing from Conceivability Arguments. Instead, my CCAs show that something is clearly going wrong within our application of Conceivability Arguments to start with, such that it cannot be the case that all of their premises are true in every application. The fact that the above conceived scenarios negate the truth of the very world from which they are posited displays that their application can bring out a contradictory state of affairs, and the reason for this fallacious conclusion must lie in the premises themselves. In this manner, we can achieve Balog's intended form of conclusion – that there is a flawed or false premise within each Conceivability Argument – and so discount consequence (i).

Let us turn to CCA1 to help explicate this reasoning further. Now, in CCA1, we had a partial zombie of the form $P \supset Q^{-1}$. This partial zombie then conceived of a further partial zombie, of the form $P \supset Q^{-2}$. Now, as noted, $P \supset Q^{-2}$ would be thought of by our partial zombie as $P \supset Q^{-1}$,

for our partial zombie does not know of his own metaphysical status as a partial zombie. In this manner, $(^{-2})$ was used rather than $(^{-1})$ only for the aid of the reader. The first argument was harnessed to utilise the metaphysical possibility of our partial zombie, and the second argument harnessed this metaphysical possibility to negate a metaphysical truth of the possible world from within which our partial zombie resides. Crucially, both arguments have mirroring premises, and so if our initial partial zombie argument is true, so too is our CCA about that zombie. However, since we have a fallacious conclusion in our CCA, it follows that one of our partial zombie's premises must be flawed, and so reductio, one of the premises in our initial partial zombie Conceivability Argument must be also.

As noted at the beginning of this subsection, I use physicalism here only as a placeholder: one could just as easily propose Counterfactual Conceivability Arguments that harness the metaphysical possibility of a world about which dualism is true, or of one about which Russellian panpsychism is true, and use the same above method to produce a contradictory conclusion which would illuminate a flawed or false premise in its corresponding initial argument.

We can see now that the consistent use of Conceivability Arguments is yielding contradictory results; Conceivability Arguments are not doing all that they are cut out to do. So far, I have entertained numerous Conceivability Arguments that negate a wide spectrum of hypotheses regarding consciousness, and also three CCAs which negate a metaphysical truth about the world from which they are posited. While the former class of Conceivability Arguments suggested that it might be the case that our entire spectrum of hypotheses are wrong, I harnessed the contradictions found within the CCAs to propose that we shouldn't be led to the conclusion that all of my previous arguments' premises are true, and that our entire spectrum of hypotheses

are incorrect. Clearly, the CCAs contain (at least) flawed premises – by virtue of their registered contradictions – and so the arguments that they are built from must too similarly. In this manner, at this point in my analysis, consequence (ii) and (iii) both remain possibilities.

At this point, it is unclear how the user of Conceivability Arguments should reconcile the contradictory state of affairs illuminated above: both the truth and the falsity of physicalism in each of our three worlds ($P \supset Q^{-1}$, $P \supset Q^{+1}$, and $P \supset Q^{xn}$) appear to be equivalently justified. However, in this subsection we have been in the privileged position of knowing the metaphysical state of affairs within our conceived worlds: it *just is the case* in those worlds that physicalism is true, and so it cannot be the case that physicalism is false. And, since we are in the privileged position of knowing what state of affairs govern those worlds, we can see the contradictions that arise from Conceivability Arguments posited from within them. Crucially, however, the entities proposing those Conceivability Arguments cannot as they do not know state of affairs governing the worlds in which they reside: our partial zombie, $\sim\text{zombie}^{+1}$, and ideal $\sim\text{zombie}$ have used Conceivability Arguments to conclude that physicalism is false in their world.

Interestingly, the clear contradictions displayed in the above Conceivability Arguments are only recognised when Conceivability Arguments are proposed from a world where the relation between P and Q is known (to us). If Conceivability Arguments work, it is the case, necessarily, that if P then Q^{-1} in our initial partial zombie world, regardless of the fact those partial zombies do not know it. What is important here, is that we know it, and so only we can see the contradiction that arises from the conclusion of their Conceivability Argument. We cannot recognise any potential contradictions resulting from our own initial argument, however, as the relation between P and Q on this world is the very question at hand.

I believe that we should extend this concern to our own application of Conceivability Arguments. Like our partial zombie, $\sim\text{zombie}^{+1}$, and ideal $\sim\text{zombie}$, we too are not in a privileged position about our metaphysical situation; this is the very reason that we proposed Conceivability Arguments in the first place. Along these lines, our own Conceivability Arguments could also be telling us things that contradict the truth about our own world. We would be unwise to then put our faith in Conceivability Arguments when considering the truth or falsity of *any* of the aforementioned hypotheses: any one of these hypotheses may well still turn out to be correct, but Conceivability Arguments are not (at least, currently) of any aid in this venture. The question that remains, is where and why our Conceivability Arguments are flawed.

4.4.2 Addressing the Issue

We have now discounted consequence (i): that all of [my previous] arguments' premises are true in each application, and our entire spectrum of hypotheses are incorrect. Accordingly, we are left with consequence (ii), or (iii): that at least one of each (of all the previous) arguments' premises are false, and so they cannot provide us with a true conclusion about any philosophy of mind hypothesis; or, that an adequate justification for the truth of at least one of their premises is currently unattainable, and so they are currently of no use within the Mind-Body debate. Both consequences necessitate that some degree of fault lies within at least one premise of each argument.

In this final subsection, I shall locate this fault primarily in P.1, specifically with regards to the claim that any of the previously stated scenarios are *ideally* conceivable. Crucially, this is not to say that the connection between conceivability and possibility is not also flawed. While

Balog proposed that her argument illustrated this fault, I believe that my Balog-inspired cases – and the analysis preceding it – are simply more indicative of a fault within our application of conceivability. I shall propose that we do not have sufficient reason in any instance concerning P or Q to make judgements about what would be known from ideal perspectives, as our attempted appeals to this ideal perspective simply amount to brute assertions that beg the question against the theory each respective Conceivability Argument is attacking. After this, I will locate a further and similar issue in the claim that the primary and secondary intensions of Q coincide. Finally, I will harness these analyses to propose that consequence (iii) is the correct conclusion to draw from my analyses.

Throughout this chapter, we have entertained rebuttals that claim opponents' Conceivability Arguments to beg the question. This potential issue was also proposed to be indicative of why some of the previous Conceivability Arguments might appear more *prima facie* palatable to the reader than others. This issue was perhaps most prevalent in the Conceivability Argument against property dualism (as motivated by the Nonphysical/Phenomenal Gap). Recall, this argument stated that:

- | | |
|-------|---|
| P.1 | It is conceivable that NP&~Q. |
| P.2 | If it is conceivable that NP&~Q, it is primarily possible that NP&~Q. |
| P.3 | If it is primarily possible that NP&~Q, then NP&~Q is secondarily possible. |
| P.4 | If it is secondarily possible that NP&~Q, it is metaphysically possible that NP&~Q, and so property dualism is false. |
| <hr/> | |
| C. | Property dualism is false. |

Now, it was noted that our property dualist might take issue with the idea of the totality of nonphysical properties not entailing phenomenal properties. Indeed, for the property dualist,

non-physical properties are supposed to be phenomenal properties, and so any account of NP should then also account for *all* phenomenal properties. I acknowledged that while this may be true, one will only find this response convincing if they first assume that phenomenal properties are nonphysical, and this has not yet been established. However, it must be noted that the converse is also true: one will only find the property dualists response *unconvincing* if they first assume that phenomenal properties are not nonphysical. In this manner, we can see how the proponent of this Conceivability Argument against property dualism appears to assume that property dualism is false in order to make his conceivability claim and rebuttal convincing. On the same lines, we saw how the physicalist can make the exact same claim as the property dualist. In order to motivate Conceivability Arguments against physicalism, one must first assume that phenomenal properties are not physical properties: to claim that $P \& \sim Q$ is conceivable, one has to first assume that Q is not P .

Although there will obviously be biases against whichever theory one is attacking, the appeal to respective epistemic gaps and ideal perspectives was proposed to extrapolate away from simple question begging to establish genuine a metaphysical proof. Indeed, this was Chalmers' entire intent behind modifying the Simple Conceivability Argument into its more complex form. However, since the claim that Conceivability Arguments beg the question persists, and since we know there must be (at least) a flawed premise within the argument, there are two conclusions that one may draw at this point: either, ideal conceivability has been achieved but is simply not a reliable guide to possibility, or, in extrapolating our initial bias into a position of 'ideal rational reflection,' we simply do not achieve the required ideal perspective needed to make possibility claims.

Clearly, we do not yet have an idealised knowledge of P and Q. However, by virtue of our incomplete epistemic position, I propose that there is also a clear sense in which we might be unable to accurately extrapolate to an idealised perspective about those variables. Indeed, the only way to attempt to achieve such a perspective is to first assume a certain relationship between the relevant variables, and then to extrapolate that same reasoning to its logical limit. However, such an extrapolation must then necessarily have built into it some initial assumption about the relevant variables and their interaction, and as such, must provide some form of reiteration of that initial biased assumption. This is the very reason that Conceivability Arguments are repeatedly claimed to beg the question; any appeal to ideal perspectives regarding current epistemic gaps simply amounts to an extrapolation of initial biases regarding whether or not those gaps will remain impassable. Accordingly, there is a clear sense from the offset in which Conceivability Arguments might simply reiterate where their proponents' loyalties already lie. The issue as it stands, is that such forms of reasoning do not tell us anything useful about the question at hand.

To help illustrate this further, let us return to CCA3, where an entity with a maximal amount of Q conceives of an entity with more Q than himself, but grounded in the same P. Recall, this argument stated that:

P.1	It is conceivable that $P \supset Q^{xn}$.
P.2	If it is conceivable that $P \supset Q^{xn}$, it is primarily possible that $P \supset Q^{xn}$.
P.3	If it is primarily possible that $P \supset Q^{xn}$, then $P \supset Q^{xn}$ is secondarily possible.
P.4	If it is secondarily possible that $P \supset Q^{xn}$, it is metaphysically possible that $P \supset Q^{xn}$, and so it is possible that there is a world about which physicalism is true.
C.1	There is a possible physicalist world in which $P \supset Q^{xn}$.

P.5	If $P \supset Q^{xn}$ is metaphysically possible, it is possible that there is some ideal $\sim$ zombie ($P \supset Q^{xn}$), who can conceive of an ideal $\sim$ zombie ($P \supset Q^{xn+}$).
P.6	If some ideal $\sim$ zombie ($P \supset Q^{xn}$) can conceive of that $P \supset Q^{xn+}$, he can produce a Conceivability Argument about $P \supset Q^{xn+}$ which shows physicalism to be false.
C.2	There is a possible physicalist world ($P \supset Q^{xn}$) in which Conceivability Arguments show physicalism to be false.

As noted previously, it might be immediately apparent to the reader that an ideal $\sim$ zombie should not be able to (ideally) conceive of this (at least if conceivability acts as a genuine guide to possibility). He is an ideal $\sim$ zombie, and so has all of the Q possible to be manifested by his P . However, he has seemingly conceived of a non-possibility where P grounds more Q than his maximal amount. Again then, there must be a fault within our ideal $\sim$ zombie's conception of $P \supset Q^{xn+}$, or, more broadly within the generally accepted link that conceivability is understood to hold with possibility.

While one might be tempted to use CCA3 to diminish the alleged link between conceivability and possibility, I believe that what is more prevalent here – assuming the validity of basic argument – is that if our ideal $\sim$ zombie had genuinely achieved a position of ideal rational reflection, he should have found $P \supset Q^{xn+}$ inconceivable from the offset. Firstly, regardless of whether the link between conceivability and possibility holds, a square circle, a married bachelor, or the statement ' $1 + 2 = 4$ ', is, by definition, inconceivable for any ideal reasoner. For this same reason, there is a clear sense in which any truly ideal perspective about those P and Q^{xn} should rule out P grounding Q^{xn+} ; Q^{xn} is maximal, such that there can never be a greater value of Q , by definition. Secondly, and even more prudently here, is that P grounds Q^{xn} by analysis in our ideal $\sim$ zombie world. Because of this, P *logically entails* what is essentially required for Q^{xn} , and so, by definition, any truly ideal reasoner should deduce that if P then Q^{xn} .

For these reasons, it appears that our ideal ~zombie simply might not have established a genuinely ideal perspective about P and Q.

Alongside those aforementioned issues surrounding the Conceivability Argument against property dualism, it seems most likely that our ideal ~zombie has simply *assumed* that the Hard Problem between P and Q will remain impassible and then extrapolated from his initial bias against physicalism into an alleged ideal perspective which can then be used to ‘prove’ that initial assumption. Crucially, since what he allegedly ideally conceived of should be ideally inconceivable, he cannot have actually ideally conceived of it in the first place. I believe that this is indicative of the fact that he simply still does not know enough about the variables to speak of anything more than his *prima facie* antecedent assumptions; our ideal ~zombie has displayed that the ‘ideal’ perspective we think we achieve in our notion of complex conceivability merely amounts to a brute assertion that begs the question against the theory at hand. Furthermore, and as noted previously, our own attempted extrapolations to this ideal perspective could be as flawed as our ideal ~zombie’s. In this manner, while ‘ideal’ conceivability may still track with possibility, in practice, it is not clear whether we actually ever achieve ideal conceivability about the relevant variables here.

Contrary to Balog’s claim that the ultimate fault lies in the link between conceivability and possibility, I propose that our particular application of ‘ideal’ conceivability in P.1 fails from the offset (Balog, 1999: 526). As noted, this is not to say that there is no further fault in the link between conceivability and possibility, but rather simply that at least our application of ‘ideal’ conceivability is flawed. Any attempt to deliberate on what can be ruled out *a priori* from an ideal perspective necessitates an extrapolation from our current *a priori* knowledge, otherwise the conceivability remains of the *prima facie* form. To achieve this ideal perspective is then to

harness current assumptions about particular epistemic gaps concerning our concepts, assume they will remain once in possession of an ideal knowledge about those concepts, and then to conclude that from that ideal perspective, such gaps will remain impassable. This clearly begs the question against whatever theory one intends to apply a Conceivability Argument against, and simply amounts to the same sort of *prima facie* conceivability claim that Chalmers hoped to deter.⁶⁶ Following this logic, the visible contradiction in my CCAs likely originates from a flawed application of what is ideally conceivable, rather than there being an implicit flaw necessarily shared amongst a premise of all Conceivability Arguments: the ideal $\sim$ zombie is certainly wrong to posit the ideal conceivability of something with a greater than maximal amount of Q, and, crucially, so too might we be wrong to posit the ideal conceivability of an ideal $\sim$ zombie in the first place.

I propose that each other Conceivability Argument remains guilty of this same fallacious extrapolation. For example, in our initial Conceivability Argument against physicalism, the proponent assumes that the Hard Problem will remain impassable, so as to prove that physicalism is fundamentally flawed; in the Conceivability Argument against Russellian panpsychism, the proponent assumes that the Combination Problem will remain impassable, so as to prove that Russellian panpsychism is fundamentally flawed; and so on. In each of these instances, one claims that a certain epistemic gap remains impassable at the ideal perspective, so as to ‘prove’ that the theory in question is fundamentally flawed by virtue of that gap. This reasoning does nothing to benefit the proponent of Conceivability Arguments, as it merely assumes the very thing its proponent sought to prove.

⁶⁶ Similar conclusions have been drawn by other authors (cf. Carruthers & Shier (2012) and Brown (2010)), but via different means. For potential criticisms of this conclusion, see Chalmers (2002), Stoljar (2006), Alter (2016), and Levin (2012).

Now, this is not to say that Chalmers' strict criteria concerning conceivability is itself necessarily flawed. Perhaps, if an ideal perspective about P and Q was genuinely achieved, it might be the case that certain Conceivability Arguments would be proven successful. However, at that point we would know whether some particular form of physicalism, dualism, or Russellian monism was true, and so we wouldn't need Conceivability Arguments to help us establish such. Along these lines, a Conceivability Argument could potentially show any one of the stated philosophy of mind hypotheses to be true or false, but only if in fact that theory was indeed true or false. Furthermore, we could only know that a Conceivability Argument has shown this iff we already knew the truth or falsity of the theory in question. Crucially then, while Chalmers' assertion that ideal, primary, negative conceivability about P and Q could still potentially be used to prove the relation between them via a Conceivability Argument, it does not follow that we can rule out the relevant relations now via a priori means. Indeed, I suspect that a posteriori means will be wholly necessary to establish these particular relations, and so that future discoveries in this regard will make appeal to Conceivability Arguments impotent.

To help reinforce this claim further, let us turn to Chalmers' claim that the primary and secondary intensions of Q coincide. Chalmers built upon Kripke's notion that the essence of a phenomenal state just is what it feels like to be in that state, i.e., its appearance properties. With this upheld, he claimed that once we have (ideally negatively) primarily conceived of $P \& \sim Q$, we can infer from its primary intension that its secondary intension is possibly true, and so that it is secondarily and so metaphysically possible. In this manner, we can bypass the typical empirical requirements for secondary conceivability, and reach it through a priori consideration about the primary intensions of Q. While 'ideal' conceivability was then necessary to produce a robust tracker to genuine possibility claims, 'secondary' conceivability was necessary to translate that claim into a genuine metaphysical possibility.

Now, while it is clearly a fair judgement to claim that the essential property of, for example, pain, *might* be just what it is like to be in pain, to claim that it *just is* what it is like to be in pain might also be guilty of question begging. First and foremost, to claim that the primary intension of Q coincides with the secondary intension of Q appears to be built from the assumption that Q cannot be defined in terms of P. At first glance, this might not seem too contentious. Pain is a qualitative concept, and so intuitively can only be understood in qualitative terms. As Kripke noted, “Pain... is not picked out by one of its accidental [appearance] properties; rather, it is picked out by the property of being pain itself, by its immediate phenomenological quality:” pain is *just a feeling* (Kripke, 1980: 15). However, while this may be *prima facie* true, we certainly cannot yet claim it to be ideally true. Indeed, to claim that Q can *only* be understood in terms of Q assumes that it cannot be, nor will come to be understood (in whole, or in part) by P. This is simply not a claim we are capable yet of making. Perhaps, the type-c physicalist will be vindicated, and Q will be explained and entailed by P in its totality, such that no epistemic gap remains. Or, indeed, perhaps not. What matters at this point, is we simply do not know enough to make ideal claims with regards to either eventuality. For the same reasoning that P.1 was flawed, so too is the assumption that the primary and secondary intensions of Q coincide.

Harnessing both issues, we can see that: (1) the Conceivability Arguments here do not satisfy the strict notion of conceivability argued to be a good approximator to (primary) possibility; and (2), since we cannot yet establish an ideal link between the primary and secondary intensions of Q, we cannot translate whatever notion of primary possibility we have about P and Q into genuine metaphysical possibility claims. Accordingly, the conclusion of each Conceivability Argument can no longer be deemed to carry the required weight to support or

attack any philosophy of mind hypothesis. In fact, all that we need is for one of these two objections to obtain in order to negate the usefulness of Conceivability Arguments. Either objection is sufficient to make the charge that all Conceivability Arguments can provide is a reinforcement of the notion that some or all of the stated philosophy of mind hypotheses *could* turn out to be false. This is not a particularly helpful reinforcement, as it is indeed just a reiteration of the question at hand: ‘how is it that Q?’

With these faults illustrated, let us turn back to our two remaining potential consequences: that (ii) at least one of each [of my previous] arguments’ premises is false, and so they cannot provide us with a true conclusion about any philosophy of mind hypothesis; or, that (iii) an adequate justification for the truth of at least one of their premises is currently unattainable, and so they are currently of no use within the Mind-Body debate. I have argued that it is unclear whether or not we can currently achieve the perspective of ideal rational reflection needed to ideally conceive of any of the stated scenarios. In this manner, regardless of whether or not Chalmers’ link between conceivability and possibility holds, we have not established an adequate justification for P.1.

Since the link between conceivability and possibility is generally well accepted throughout the philosophical tradition, I proposed that a Conceivability Argument could potentially come to prove some certain hypothesis true if that hypothesis was in fact true. However, I have argued that we would only know which Conceivability Argument is successful if we already knew the truth or falsity of the hypothesis in question. Accordingly, since I have not explicitly discovered any necessarily false premises within the argument, we can say that it would be unwise for us to currently accept consequence (ii).

Now, it could turn out that one or more of the previous arguments is in fact true (even if consequence (iii) holds), and so (ii) could be discounted. However, assuming the general link between conceivability and possibility, since we do not yet have an ideal perspective, we do not already know which epistemic issues will be explained away or which hypothesis is or is not true. Accordingly, we cannot yet know which, if any, of these above arguments is successful. In this light, it seems that consequence (iii) is the most appropriate for us to currently uphold: we simply do not have an adequate justification for the truth of at least one of the premises in each argument, and so Conceivability Arguments are currently of no use within the debate. Furthermore, since Conceivability Arguments may only be informative about a hypothesis iff it is in fact true or false (and we already know that fact), it is unclear whether they ever will be of any use further down the line.

I would recommend, therefore, that Conceivability Arguments are relinquished from the debate altogether. As we have seen, they can be used to negate any philosophy of mind hypothesis and to produce contradictory conclusions, and this is all due to their being plagued with a pervasive question begging from the offset. I have proposed that we do not currently have an adequate justification for the truth of their premises, and so they are currently of no use within the debate. Furthermore, since I have shown that they will only be informative iff we already know the truth or falsity of the theory in question, it is unclear whether they will ever have any use other than to bolster pre-existing biases.

For the Russellian panpsychist, this recommendation has the unfortunate consequence of removing one of the most prevalently appealed to arguments for their theory. However, its omission from their repertoire is certainly worthwhile. As we have seen, Conceivability Arguments appear to threaten Russellian panpsychism to the same degree as physicalism, and

so it serves them no better than their rivals. However, I have shown that all stated applications of the argument are fundamentally flawed, and so they cannot provide a compelling case for or against any theory. As such, its removal necessitates that a new, improved argument must be levelled for Russellian panpsychism in order to establish its theoretical promise over the alternatives. Indeed, with the Combination Problem still looming, providing such an argumentative brace is now wholly necessary for the theory to be taken seriously. With this in mind, the remaining chapter will formulate a new argument for Russellian panpsychism. By not restricting myself to merely a priori reasoning, and through avoiding appeals to Conceivability Arguments altogether, I shall formulate a new argument that should positively distinguish Russellian panpsychism from physicalism as the most theoretically promising solution to the Mind-Body Problem.

5

An Evolutionary Argument Against Physicalism

Physicalism's Reduction into Panpsychism

This chapter will propose a new argument against physicalism and for Russellian panpsychism. Building from the work of William James (1890) and William Clifford (1874), I will propose a new 'Argument from Evolution,' which shows how physicalism struggles to account for consciousness within the evolutionary picture. My argument begins with two core claims: (1) that the physicalist *should* endorse the precision of consciousness, and (2) that the physicalist who endorses the precision of consciousness will endorse that consciousness is a result of saltation, as opposed to being something fundamental. Upholding these claims, and by working down our evolutionary trajectory with regards to our evolutionary relatives, I will show how the physicalist cannot find any non-arbitrary point where consciousness could have originated aside from the origin of the first living things. I will then show how this conclusion serves only to worsen the epistemic issues already suffered by physicalism, such that it can no longer be taken to be preferable to Russellian panpsychism.

This chapter will also serve two functions. First, it will serve as a safeguard to Russellian panpsychism: even with the Combination Problem looming, Russellian panpsychism can still provide a more attractive theory of consciousness than physicalism, provided one avoids attempting to motivate it with Conceivability Arguments. Second, I shall propose that potential responses to the Combination Problem can be built from the logic of my argument, and so this

chapter can serve as a foundation for later responses. I shall briefly investigate three potential avenues for the Russellian panpsychist in this regard. After this, I will introduce some historical context with regards to predominant developers of evolution that helps to furnish the conclusion of this chapter. With this in mind, I shall provide Russellian panpsychism with a theoretical ‘safety net’ over physicalism with regards to establishing its internal coherence, and also the grounding to resolve its own internal issues.

5.1 ~ The Problems that Remain

Before delving into my new argument against physicalism and for Russellian panpsychism, we must take a moment to observe the philosophical landscape for both theories now that Conceivability Arguments have been relinquished. As we saw in the last chapter, the existing Conceivability Arguments within the Mind-Body debate are plagued by a pervasive issue: they do not have adequate justification for the first premise of their argument and so are currently of no use within the debate. Now, while I have recommended the removal of Conceivability Arguments from the debate, the epistemic issues suffered by each theory that were harnessed into P.1 of their respective Conceivability Arguments still remain. With this in mind, this subsection will revisit these existing problems, before the later subsections will begin to formulate arguments around those suffered by physicalism.

Although Conceivability Arguments traditionally form a popular argument in support of Russellian panpsychism and against physicalism, as we saw in Section 4.2.2, we can harness issues associated with Russellian panpsychism’s Combination Problem into an analogous Conceivability Argument against it. If we also recall from Section 2.3.2, the Combination Problem exists *independently* of Conceivability Arguments and so does not need Conceivability Arguments to motivate its existence. As we saw there, in its most general form,

the Combination Problem is the problem of accounting for how micro-experiential quiddities combine to make macro-experience. In particular, we saw how the Subject Summing Problem, and the Boundary Problem illustrate (at least) epistemic problems for Russellian panpsychism, which might turn out to be ontological ones. Even without Conceivability Arguments then, these problems still present a robust challenge to Russellian panpsychism's internal coherency.

Like Russellian panpsychism, while my removal of Conceivability Arguments does also remove a popular argument against physicalism, the epistemic issue suffered by physicalism – namely, the Hard Problem – that was harnessed into the Complex Conceivability Argument still remains.⁶⁷ As noted in Section 1.3, 'physicalism' has been used as a placeholder for both its type-b and type-c variants, both of which recognise the Hard Problem as a current (and at least) epistemic issue suffered by their theory. As we saw there, this issue also exists *independently* of Conceivability Arguments, such that it remains an issue the physicalist must still come to resolve, even after Conceivability Arguments have been relinquished.⁶⁸ While Conceivability Arguments are often used to help illustrate the Hard Problem, the Hard Problem in-and-of-itself is simply the problem of explaining why and how phenomenal experience arises from the physical processes in the brain.

If we recall, type-b physicalists accept that the Hard Problem is an epistemic gap that will not come to be resolved but deny that the problem is an ontological one in nature. In this manner, p-zombies were conceivable for the type-b physicalist, but they were allegedly not metaphysically possible. We can say then that for the type-b physicalist, the *concept* of consciousness is distinct from typical physical concepts but that phenomenal states can be

⁶⁷ I say 'epistemic' rather than 'ontological' here, as without Conceivability Arguments, the opponent of physicalism must find a new way to defend this epistemic gap being ontological in nature.

⁶⁸ The type-b physicalist will still maintain that the Hard Problem exists, even though they don't think it is something that we will come to explain away, nor necessarily need to.

identified with physical states, nonetheless. Crucially, however, this identity is one which is *primitive*, such that it is not deducible *even* with a complete physical understanding of the world. Type-c physicalists, on the other hand, accept that the hard problem is a currently robust epistemic gap, but argue that it will eventually come to be closed as we uncover more about the mysteries of consciousness and its relation to the physical. In this manner, p-zombies are (prima facie) conceivable to the type-c physicalist *now*, but not ideally conceivable. We can say then that for the type-c physicalist, while there appears to be a robust epistemic gap separating the physical from the phenomenal, this gap might come to be closable in principle, perhaps through the development of more complex theories, or perhaps by some more conceptually advanced inquirer (Chalmers, 2002: 255-60).

My removal of Conceivability Arguments from the debate does remove a popular tool against physicalism: for both variants at hand, the current existence of the Hard Problem was harnessed into P.1 so as to establish an alleged ontological conclusion. However, as with Russellian panpsychism, the removal of the tool does not remove the problem that the tool hoped to advance. For both type-b and type-c physicalism, the Hard Problem (in at least its epistemic form) still remains. All that has changed in this regard is that there is no longer an argument at hand which allegedly shows this epistemic problem to be an ontological one. Whether any of these epistemic gaps present a genuine worry to these theories remains to be established.

Thus, we return to the theoretical ‘level playing field’ between Russellian panpsychism and physicalism that was finalised in Chapter 2. Both have respective (and unresolved) epistemic issues, and so in order to give one or the other theory the advantage, progress must be made into explaining one of the problems away, or a new argument which worsens the problem of the other theory is needed. In this regard, each theory must show how their respective problem

is not a problem at all, or that the other's problem is *more* of a problem than their own. With this in mind, in this chapter I will propose an Argument from Evolution that worsens the respective epistemic issues of type-b and type-c physicalism and expound how the result of this is much more damaging to physicalism than the Combination Problem is for Russellian panpsychism. In doing so, I provide Russellian panpsychism with a 'safety net,' such that even with the Combination Problem unresolved, it can still be theoretically preferred over physicalism.

5.2 ~ The Traditional Argument from Evolution.

Arguments of a similar vein to that which I will propose in this chapter have already been suggested by William Clifford (1874) and William James (1890). Known now as 'Continuity Arguments,' these authors looked to evolution in the hope that it would help to bridge the gap that divides the physical and the mental sciences (Clifford, 1874: 36; Roelofs, 2019: 18). Now, clearly anyone who is inclined to take consciousness seriously will intuitively be drawn to the question of how to fit consciousness within the evolutionary picture: we are creatures that have evolved, and we are creatures that are conscious. In this manner, consciousness, and the evolution of it must then fit somehow into our evolutionary trajectory.

Per Clifford and James, for the physicalist to fit our consciousness into the evolutionary picture, they must be theoretically able to point to a time in which consciousness began accompanying the physical processes which allegedly necessitate or identify with it, and from which, continued to evolve into the state in which we enjoy it today. However, there is arguably no clear time in our evolutionary trajectory that one could ever point to where this should occur: our development was gradual and incremental at every stage. In this manner, it appears impossible to point out some particular line of descent where our physical complexity became

sufficient so as to be accompanied by an “irruption” of consciousness (James, 1890: 148; Clifford, 1874: 61). As Clifford puts it, we cannot suppose that “so enormous a jump from one creature to another should have occurred at any point in the process of evolution as the introduction of a fact entirely different and absolutely separate from the physical fact” (Clifford, 1874: 61). The pervasive question at hand is then, if there is no clear evolutionary ‘leap’ from a nonconscious living thing to a conscious living thing within our evolutionary trajectory, then where should we point to for its initial occurrence?

Both Clifford and James propose that since the physicalist will fail to find some non-arbitrary point within our development that is significant to the existence of consciousness, then consciousness must have been present at the origin of living things (James, 1890: 152). Grounding this reasoning is the assumption that evolution is only ever a continuous process: one which ‘stacks’ pre-existing properties into more complex forms. Following this reasoning, and if we are to accept the evolutionary model, then even the very lowest of organisms must have something that it is like to be. For example, even the amoeba that swims around in our blood must be endowed with consciousness. While this might be of an inconceivably simple amount when compared to us, it must at least share the very same nature as our own consciousness if consciousness is to have been passed onto us (Clifford, 1874: 61).

More formally, we can put James’ and Clifford’s argument as follows:

- A. Evolution can only assemble pre-existing properties into further complex forms.
- B. Given A, evolution can only assemble pre-existing consciousness properties into further complex forms.
- C. Consciousness must have been present at the origin of living things.

I shall use this initial argument by Clifford and James as a starting point for my own Argument from Evolution. However, my argument will digress in a number of important ways. First and foremost, I believe that there is a sense in which their argument might over-generalise evolution from the offset. Indeed, it seems intuitive that evolution can, in many ways, produce ‘new’ properties that were not present at the origin of living things. For example, the fact that humans and other animals have the property of sight does not entail that the very first living things did also. In this sense, without further argumentation and development, it is not clear whether these authors can achieve their intended conclusion. Second, while both authors are swift to treat their considerations as an implicit argument for panpsychism, as we shall see in the next section, if one is to establish their intended conclusion one must first establish that consciousness is ‘precise’.⁶⁹ My Argument from Evolution will build from and avoid these issues, so as to present a robust argument against physicalism, and for Russellian panpsychism.

5.3 ~ A New Argument from Evolution Against Physicalism

This subsection will propose my novel Argument from Evolution against physicalism. Building from the work of William Clifford (1874), and William James (1890), I will structure an argument which discredits the claims of type-b and type-c physicalism. My argument will develop two claims: (1) that the physicalist should seek to endorse the ‘precision’ of consciousness, and (2) that the physicalist who endorses the precision of consciousness will endorse that consciousness is a result of saltation, as opposed to being something fundamental.⁷⁰ Upholding these claims, and by working down our evolutionary trajectory with regards to our evolutionary relatives, I will show that there is no relevant point at which the physicalist can point to for consciousness to ‘wink in’ (Chalmers, 1996: 297). Instead, I will

⁶⁹ As shall be defined shortly, a property is ‘precise’ if it does not admit of borderline cases.

⁷⁰ As shall be defined later, ‘saltation,’ refers to a sudden and large mutational change.

show how the physicalist searching for a non-arbitrary point for consciousness to begin must endorse ‘biopsychism,’ but that this position serves only to worsen the epistemic issues already suffered by physicalism. In particular, I will then show how this position threatens either:

- Physicalism’s reduction into Russellian monism,
- Its inheritance of the Combination Problem,
- Its worsening of the Hard Problem,
- Or indeed, some combination of all of the above.

In any of these instances, physicalism must be taken to be less theoretically promising than Russellian monism, all other things considered.

The claim that physicalism is incompatible with precise consciousness is not new within the literature. For example, Michael Antony (2006: 515-38) and Joshua O’Rourke (2018) argue that if consciousness is precise, then many particular forms of physicalism must be false. Similarly, Philip Goff (2014), and Michael Tye (2021), argue that the precision of consciousness would support panpsychism, or a form of representationalist panpsychism, respectively. This section will present a novel argument of similar conclusion. Crucially, while the conclusion may be similar, the content, structure, and, indeed, goal of my argument will differ.⁷¹

⁷¹ I shall lay out these extant arguments in more detail, once the relevant terminology has been set up.

5.3.1 The Precision of Conscious Experience

As noted, much of my argument in these forthcoming sections will be reliant upon the claim that consciousness is *precise*, such that it does not admit of borderline cases. In this manner, my later argument will take the form of: *if* consciousness is precise, *then* then the physicalist will fall victim to the forthcoming issues with regards to evolution. With this in mind, this subsection will highlight the particular distinctions and variables relevant to my later argument. Since I intend for my later argument to extend beyond a standard incompatibility thesis into an argument against physicalism simpliciter, I will also present an argument for why the physicalist should seek to endorse the claim that consciousness is not vague, and, consequently, why they will likely find my subsequent argument a genuine challenge to their thesis.

At the most surface level, a property is vague if it admits of indeterminate, or borderline cases. Within our natural languages, vagueness is allegedly a common state, such that most descriptors or properties are vague (Hall, 2023: 682). For example, ‘baldness’ is allegedly a vague property; there are many bald people, many people who are not bald, and also a vast spectrum of people in between with varying degrees of hair loss. Within this spectrum, there is allegedly no clear point at which one can say, “that is the place, there!” where non-baldness turns into baldness. Instead, these midground instances are neither determinately bald nor determinately non-bald, in such a manner that the property ‘bald’ is vague (Field, 1994: 410). With this in mind, to say that consciousness is vague is to say that consciousness admits of indeterminate, or borderline cases.

A property is precise, on the other hand, when it does *not* admit of indeterminate, or borderline cases. For example, the property ‘squareness’ is precise; either a thing has the property of having four equal sides, and four equal angles, or it does not. Likewise, the property of being

a prime number is precise as it is not possible that something is such that it is borderline whether it is prime (Hall, 2023: 670). For such non-vague properties, we can say that there is a clear ‘line in the sand’ between things that possess that property, and those things that do not. With this in mind, to say that consciousness is precise is then to say that “there is either a light on inside, or there is not,” or that “there is either something it is like to be a thing, or there is not;” there can be no borderline cases such that something can be vaguely conscious or not.

Within contemporary literature, there are three dominant accounts of what it means for a thing to be vague, or, more precisely, why certain properties *are* or *appear to be* vague. The first account of vagueness is ‘metaphysical’. Now, there is vagueness of this variety if for some object and some property, there is genuinely no determinate fact as to whether or not that object exemplifies that property (Merricks, 2001: 145). In this manner, metaphysical vagueness suggests that certain features or properties of the world are inherently indeterminate in-and-of-themselves, such that the vagueness corresponds to the nature of reality itself (van Inwagen, 1990: 228ff; Tye, 1990: 583). If the property of baldness is metaphysically vague, one might say that there is no determinate fact of the matter as to whether a questionably bald man exemplifies the property of being bald, such that the corresponding proposition that someone exemplifies the property of baldness is neither true nor false.

The second account of vagueness is ‘epistemic.’ This form of vagueness alleges that indeterminacy arises from our limited knowledge of the world, rather than from the nature of the world itself. This account suggests that certain propositions or concepts lack clear-cut boundaries because we do not have the cognitive or empirical means to yet determine them with precision; epistemic vagueness is explained completely by and is nothing over and above our not knowing some relevant fact or facts (Merricks, 2001: 146). In the case of baldness, the

alleged vagueness would be explained by our not knowing enough about the nature of baldness, or the propositions we express about it in any given context. In this manner, once we locate our relevant ignorance, the vagueness would allegedly dissipate (Williamson, 1994: 12ff; Sorensen, 1998).

Finally, the dominant account of vagueness is ‘linguistic.’ This account proposes that vagueness is never metaphysical nor epistemic but is instead a product of the indeterminacy or imprecision inherent in language itself (Strawson, 1964; Wittgenstein, 1973; Goff, 2014: 81). Language is inherently flexible, and so terms, expressions, and concepts denoting properties often do not have clear-cut boundaries in their application. As a result, we often use words that have broad or imprecise meanings, allowing for multiple interpretations in various contexts (Lewis, 1986: 68; DeRose, 1995: 553). In the case of baldness again, if baldness is linguistically vague, the apparent indeterminacy in the property results from the indeterminacy of the word ‘bald’. This indeterminacy would be explained by the fact that we perhaps never decided exactly what ‘bald’ should mean, or in terms of the incompleteness of the term ‘bald’ (Merricks, 2001: 146).

As noted previously, the claim that the precision of consciousness is incompatible with physicalism, or that it supports panpsychism, is not new within the literature. For example, Michael Antony proposes that there are intuitive reasons to believe that our concepts regarding conscious states are linguistically precise rather than linguistically vague, and that if these concepts are correct (at least in their precision), then many popular forms of physicalism are false (Antony, 2006: 515). For example, he proposes that one must not confuse faint, hazy, incoherent, or fleeting experiences with genuinely borderline experiences; the former have characteristic, rich phenomenologies, and so are still always determinately conscious, such that our concepts regarding them remain precise (ibid: 516). In contrast, Antony argues that our

concepts regarding neurophysiological states are vague due to their complexity. By virtue of this, Anthony argues that forms of physicalism, such as identity theory, cannot identify conscious states with specific neurophysiological states, for they possess different properties (ibid: 522).

In a slightly different vein of argumentation, Michael Tye proposes that physicalism typically entails the linguistic vagueness of consciousness, as all of the concepts available to the physicalist are vague (Tye, 2021: 6). However, he proposes that there are no possible borderline cases of consciousness, and so formulates the following potential paradox for the physicalist:

- P.1 It is an instance of the logical law of the excluded middle that either consciousness is precise, or it is vague.
- P.2 If physicalism is true, then consciousness is vague.
- P.3 If consciousness is vague, there are possible borderline cases of consciousness.
- P.4 There are no possible borderline cases of consciousness.
- P.5 Consciousness is not vague.
- C. If physicalism is true, consciousness is both vague and not vague.

(ibid: 73).

Tye moves on to assess Russellian monism as a potential resolution to this paradox, and while he finds it insufficient in-and-of-itself, he proposes a hybrid position that combines an important aspect of panpsychism with representationalism. In particular, he proposes that the way out of the paradox is to accept that various aspects of consciousness are vague, and that these should be accounted for within a representationalist view. However, he proposes that a central element of consciousness, viz. experience itself, is precise and exists as a fundamental property that lies outside the domain of microphysics (ibid: 99).

Finally, Philip Goff has argued that if property dualists are committed to the linguistic theory of vagueness, they are obliged to hold that it can never be vague as to whether or not a given thing is conscious. Upholding this and the commonsense assumption that not all macro-level objects are conscious, he argues that this entails that the psycho-physical laws governing consciousness production are precise. However, Goff finds this conclusion entirely implausible, leading him to suggest that the property dualist should reject the commonsense assumption, and instead uphold that all macro-level objects are conscious. Along these lines, Goff proposes that the property dualist who upholds the linguistic theory of vagueness will be led to a distinctly panpsychist conclusion (Goff, 2014: 75-92).

While the claim that the precision of consciousness is incompatible with physicalism or supports panpsychism, is not then new within the literature, my argument will digress from the extant argumentation in a number of significant respects. First, I shall argue that the physicalist should endorse that the concept ‘consciousness’ does *not* admit of borderline cases, and so it appears to be linguistically precise. I shall propose that, by virtue of finding no positive characterisation of borderline consciousness, nor any potential sharpening to the predicate ‘is conscious,’ one is naturally led to the conclusion that consciousness is unlike other linguistically vague properties such as ‘baldness’ or ‘heap.’ Instead, consciousness intuitively appears to be precise, in the same way that ‘squareness,’ or ‘being a prime number’ is; it does not admit of any indeterminate cases. After this, I will explain how an epistemic understanding of vague consciousness also entails a definite boundary between things that are conscious and things that aren’t.

After this, I will explore a ‘backup’ argument that considers the consequences that physicalism may face if they continue to deny the precision of consciousness. I will propose a Sorites paradox that potentially shows any form of vague consciousness to be incoherent, such that physicalism cannot then satisfy the Consciousness Constraint or might reduce into something similar to Russellian panpsychism. Finally, I shall appeal to evolution to show how the physicalist who endorses the precision of consciousness will, nevertheless, run into a number of pervasive issues that weaken its claim further and provide explicit support for Russellian panpsychism.

To say that consciousness, or the concept ‘consciousness’ is vague is to suggest that there are not clear, well-defined limits or conditions that make something conscious or not conscious; the concept consciousness, or indeed, consciousness itself, would not have a precise, fixed nature that allows for a sharp distinction between conscious and non-conscious states, or between conscious, and non-conscious entities. I shall argue that the physicalist should resist the claim that consciousness is ever vague, and instead endorse that consciousness is *precise*, such that both the concept, and the thing itself admits of no indeterminate cases, nor is inherently indeterminate in-and-of-itself.

With these clarifications in place, let us explore the reasons for why the physicalist should first resist the claim that the concept ‘consciousness’ is vague. Now, when referring to vagueness, most authors uphold the necessity of providing a positive characterisation of borderline cases, viz. a way of filling in the details that will show a sufficiently competent speaker what makes such and such an entity a borderline case with regards to a certain property (Simon, 2017:

2106).⁷² With this in mind, and so as to sufficiently ground my later argument against physicalism, I shall present a novel argument which proposes that the physicalist is unable provide a positive characterisation for vague consciousness, and so should accept that the concept ‘consciousness,’ and in turn, consciousness itself, is always precise.

Now, when we reflect on our consciousness, we can see that cognition or cognitive sophistication might admit of degrees. For example, a person endowed with more concepts or knowledge might view the world in a more enriched fashion than those who do not. However, this should be of no worry here; as previously discussed in Chapter 1, those higher order forms of cognition are not what is of interest to this thesis. The focus here is and always has been phenomenal experience: having an inner life such that there is something it is like to be you. While higher order forms of cognition might indeed admit of degree – and so may be potentially open to claims of indeterminacy – one might find that there is a clear intuitive sense in which phenomenal experience does not: you either have an inner life (which can be more or less complex), or you do not (Goff, 2014: 82).

A useful analogy here might be to consider consciousness as a windowless room with varying levels of furnishings, and a single lightbulb hanging from the ceiling. For some, the room is furnished with tables, chairs, paintings, books, and all those things that are normally – or could possibly be – housed within a room. For others, the room might be more sparsely decorated, with just, for example, a chair and a desk. Finally, for others, the room might be empty of any furnishings at all. Take this spectrum of furnishings across rooms to be the spectrum of higher order forms of cognition and take the room itself to be brains or brains and bodies, or whatever

⁷² See Antony (2006), Simon (2017), Eklund (2005), Greenough (2003), Schiffer (2003), Smith (2005), Weatherson (2003), Wright (2001), for attempts to positively characterise linguistically vague borderline cases, and see Schwitzgebel (2023), Dennett (2005), Tye (2000), and Prinz (2012) for arguments regarding the metaphysical vagueness of consciousness.

vessel it is from which consciousness occurs. Clearly, one can have more or less things in one's room: one's experience itself can be richer or poorer than another's. Crucially, however, and regardless of how dim or bright the light is in the room, the lightbulb can only be determinately on or determinately off. We can say that just like the room's size and furnishings can admit of degree, so too can the *content* of consciousness and the vessel in which it occurs. However, just like the light in the room is either on or off, so too must there be either something for which it is like to be that thing with that level of content, or nothing for which it is like to be that thing.

The claim that the content of one's experience can admit of degree should be uncontentious to the reader. However, one might respond that the analogy is flawed due to its reference of experience simpliciter as a lightbulb. In the same way that lights can be more or less dim, one might respond that so too can experience simpliciter. For example, perhaps the strength of one's phenomenal experience (or brightness of the lightbulb) is entirely dependent upon the *content* of one's experience (or the level of furnishings in the room). And, if the content of experience is tied in this way to experience simpliciter, then experience simpliciter might admit of degree in a similar sense to cognitive sophistication.

This potential objection should not worry the reader here within the parameters of my forthcoming analysis. As noted above, if there is a light on in the room, regardless of how dim that light is, the light is determinately on. Similarly, if an entity has experience simpliciter, regardless of how 'dim' that experience is, the entity appears to have something it is like to be it (Simon, 2017: 2; McGinn, 1996; Papineau, 1993: 125).⁷³ Accordingly, regardless of whether

⁷³ For analogy, we can understand light as a non-vague concept: wherever there is a photon (or many photons), there is light, no matter how dim it may appear.

the content of experience is tied to the level of experience simpliciter, a ‘dim light’ is still an inner life in the sense relevant to this thesis; there is still something rather than nothing that it is like to be that thing. In this manner, the distinction between something-it-is-like-to-be | nothing-it-is-like-to-be (or light on | light off) does not admit of degree in such a manner that the property becomes vague in a relevant sense.

To make this claim more vivid, we can ask what it would really mean for consciousness to be vague. Recall, vague properties allegedly include things like the colour red: there are colours that are determinately red, colours that are determinately not red, and there are also colours that fall into an indeterminate category between the two. If it is the case that consciousness is vague, then we should be able to provide a positive characterisation of borderline cases in the same way as we can with ‘red’, viz. a way of filling in the details that will show a sufficiently competent speaker what makes such and such an entity a borderline case with regards to consciousness.

When we reflect on our own consciousness, we can find no obvious forms of indeterminacy: even those special experiences which seem ‘half-conscious’ – such as when waking from sleep – are, upon reflection, unequivocally conscious (Papineau, 1993: 125; Tye, 2021: 13). For example, for Tye, waking is a deterministic transition that simply reflects underlying physical processes in the brain, alongside the introduction of sensory inputs (Tye, 2000: 155). In this manner, when waking from sleep we are arguably just undergoing a regulated process from a state of lowered conscious awareness to one of full conscious awareness. However, this is not a case of an indeterministic shift in consciousness, as while it may be indeterminate as to *what* you are experiencing in such instances, it is never indeterminate as to *whether* you are experiencing (Tye, 2021: 14)

While it may be unclear whether we can find any cases of indeterminacy in our own conscious experience, one may still wonder whether it might be indeterminate for some other entity, such that consciousness – broadly speaking – can be vague. Let us for the moment then suppose that consciousness can be indeterminate in this way, such that the hierarchy of entities in existence thus fall into the following *theoretical* demarcations:

Non-Conscious | Vaguely Conscious | Conscious

While this trifurcation would, if true, form a large spectrum with candidates across each category, we can point to some potential candidates within each demarcation, nonetheless. For example, we should fall within the conscious category, snails might fall within the vaguely conscious category, and rocks and pebbles would surely fall into the non-conscious category. In this manner, while we enjoy a rich inner life, perhaps there is only something vaguely that it is like to be a snail, while there is nothing it is like to be a rock or pebble.

But what could it mean for there to only be vaguely something it is like to be a thing? We are not saying that the snail has nothing it is like to be. Rather, to put the snail in the vague category is to say that that there is more than nothing that it is like to be a snail, but just not quite something that it is like to be: it is more than definitely non-conscious, but less than definitely conscious. This is just to say that the question, “What is it like to be a snail?” lacks any determinate, or definite, answer (Hall, 2023: 671). However, one might propose that herein lies the issue with this form of reasoning: if there is not nothing it is like to be a snail, then there is a clear sense in which there is surely *something* it is like to be it, however simple that something is.

Now, in virtue of ‘baldness’ being a vague property, a person can instantiate a midground instance whereby he is neither determinately bald nor determinately non-bald. For this reason, the statement “if it is not the case that Philip Goff is not bald, then he must be bald” does not appear to be true. Intuitively, the same cannot be said when we consider consciousness. From the perspective of the snail – of which it is allegedly vague as to whether there is something it is like to be – there is either something it is like to be it, or there is not. In this manner, there is not a semi-something it is like to be a thing: a semi-something *just is* a something. Perhaps, those who would assert otherwise within the parameters of this discussion are guilty of confusing the *content* or *complexity* of experience, with experience *simpliciter*. Intuitively, if the snail has anything at all of which it is like to be, it has experience simpliciter, and so it does not fall into a vague category of experience. Similarly, if the snail does not have anything at all of which it is like to be, it does not have experience simpliciter, and so does not fall into a vague category of experience. In this manner, consciousness appears to be entirely different to traditionally vague properties such as baldness or heaps.

It is a consequence of classical logical and semantic precision that everything is either square or not square (Hall, 2023: 671). The same sort of statement, however, cannot be made for vague properties or predicates. For example, it is not the case that everything is either bald or not bald. In the case of baldness, this vagueness likely stems from semantic indecision. According to the linguistic theory of vagueness, for any vague predicate, there are multiple potential ‘sharpenings’ of the predicate, such that the meaning of the predicate does not settle on any of these. Consider again the vague predicate ‘is bald.’ We might choose to stipulate that anyone whose hair covers less than 25% of their hair is bald, and that anyone with more than this is not bald. Alternatively, we could stipulate that anyone whose hair covers less than 10% of their hair is bald, and that anyone with more than this is not bald. This is an alternative ‘sharpening,’

viz. an alternative way of making the predicate precise. In this manner, we can say that ‘is bald’ is vague, because no one has bothered to single out one of its sharpening’s as the unique meaning of the predicate (Goff, 2014: 81).

As elucidated above, the predicate ‘is conscious’ – and hence the concept it expresses – does not seem to be associated with a spectrum of sharpenings and so does not appear to be vague in the same linguistic sense. While it is clear then that we can make sense of indeterminate baldness, the same cannot be said for indeterminate consciousness. There are no range of cases over which it’s indeterminate where the concept applies, nor any potential spectrum of sharpenings available so as to make the concept ‘consciousness’ more precise. Indeed, the only conceivable way of having such a range is with a functional definition of consciousness, such that there are slightly differences in the relevant functioning to form the range of cases. However, as emphasised in Chapter 1, we are not interested in type-a accounts of physicalism here, and so this option is not available to the physicalist.

For these sorts of reasons, some commentators move to a more epistemic understanding of vagueness, and argue that any potential ‘vaguely conscious’ category simply represents an issue in *our* classification of entities, as opposed to those entities being truly vaguely conscious.⁷⁴ In the case of the snail, perhaps we don’t know with any reasonable certainty if the snail has any experience simpliciter or not, and it is this issue in the classification of other minds that pushes us to classify it within the ‘vague’ category. Within this process, one looks to the snail’s behaviour and physical complexity so as to speculate that it doesn’t warrant the label of ‘determinately conscious,’ nor of ‘determinately nonconscious,’ and so that they must classify

⁷⁴ See Antony (2006), Simon (2017), Eklund (2005), Greenough (2003), Schiffer (2003), Smith (2005), Weatherson (2003), and Wright (2001) for more details.

it somewhere in between. Crucially, however, our speculation about the snail's possession of consciousness is irrelevant to the actuality of that snail being conscious or not. Following the above reasoning, it must be the case that either the snail has something it is like to be (however simple), or that it does not; the fact that we don't have a definitive answer to the question of "what it is like to be a snail?" is in no way germane to the issue of whether the snail is vaguely conscious (Tye, 2021; 13). In this manner, those that move to an epistemic reading of vague consciousness must still endorse a definite boundary in principle between things that are conscious and things that aren't; things like snails are arguably never actually vaguely consciousness.

Eric Schwitzgebel has characterised views similar to that which I present here as the *Luminous Penny View*, where having consciousness is described like having money: we might have a little money, or we might have a lot of money, but having any money at all is discretely different from not having a single cent. In this manner, the snail might have a single penny, and I might have a few thousand, and in the case of the snail, that pennysworth of consciousness is the difference between the 'lights being on', and the 'lights being off.' However, Schwitzgebel proposes that this view runs into issue when it attempts to locate the lower bound of consciousness, viz. those things with only single pennies (Schwitzgebel, 2023: 18).

Schwitzgebel proposes that those things which only have a single penny must have such an incredibly impoverished state of consciousness that they are almost empirically indistinguishable from lacking consciousness. By analogy, having only a single penny is almost empirically indistinguishable from complete bankruptcy (ibid). He asks us to consider what kind of entity would only have a pennysworth of consciousness. Things like snails, for example, have fairly complex olfactory and tactile capacities by virtue of the feelers. If they

are conscious at all, it would be natural to suppose that this consciousness would be tied to the complexity of these capacities, and thus that the snail must have more than the most minimal speck of consciousness. By consequence, to locate the minimally conscious entity, we would have to look to simpler organisms like, perhaps, worms. However, Schwitzgebel proposes that the argument again repeats itself: these entities have more than minimal sensory capacities, and so if they are conscious, then it's reasonable to suppose they have more than a pennysworth of consciousness. In this manner, we will be naturally led to the following dilemma: either the simplest possible consciousness requires the simplest possible sensory system, or we postulate some middle-complexity organism that possesses the minimal amount of consciousness despite having a wealth of sensory sensitivity (ibid: 19-20).

Schwitzgebel's explicit worry with this dilemma is that the former option seems to reduce into something similar to panpsychism, and the latter option simply appears too unlikely and unintuitive. In this manner, he proposes that we would do better to endorse the vagueness of consciousness, such that there is a gradual fade-in into something much richer than a penny (ibid: 20). Contra to Schwitzgebel, I propose that the former view is not distinctly panpsychist; positing the simplest possible consciousness for those things with the simplest possible sensory system does not in-and-of-itself entail that the fundamental constituents of the universe have experience. In this manner, it should not yet worry the physicalist here. Instead, it appears more likely that Schwitzgebel is simply uncomfortable with attributing simple experience to simple living organisms, or that he wishes to reject panpsychism (or theories similar) on the basis of his own basic intuitions. Similarly, while the latter view may seem unintuitive, it may yet still turn out to be the case. Indeed, the whole reason that the Mind-Body Problem persists is that we simply do not yet know how it is that mind and body relate, nor then do we know in what sense things are or are not conscious. Accordingly, neither horn of the dilemma provides

sufficient reasoning for the physicalist to discount the precision of consciousness. Furthermore, and as I shall present below, these options offer a much more preferential alternative to the physicalist than that which they may run into if they continue to maintain that consciousness is vague.

We have seen some intuitive reasons to suggest that the concept ‘consciousness’ is dissimilar to other linguistically vague properties, such as ‘bald,’ and more similar to precise properties, such as ‘squareness.’ Similarly, we have seen how if one moves to a more epistemic understanding of vague consciousness, they must still endorse a determinable and precise boundary in principle between conscious and non-conscious things. At this point, the physicalist may opt simply to propose that although there are vague concepts and properties in the world, consciousness is not one of them. However, so as to cover all bases, let us now consider a of ‘backup argument’ for this conclusion, that considers some consequences that physicalism might run into should they continue at this point to deny the precision of consciousness. Let us return to our theoretical trifurcation above, and the token entities attributed to each demarcation:

Non-Conscious		Vaguely Conscious		Conscious
Rocks and Pebbles		Snails		Humans

Now, for this trifurcation to hold, it must be the case that there is a category of entities that are determinately non-conscious, a category of entities that are determinately conscious, and some third category of entities that are neither determinately conscious nor determinately non-

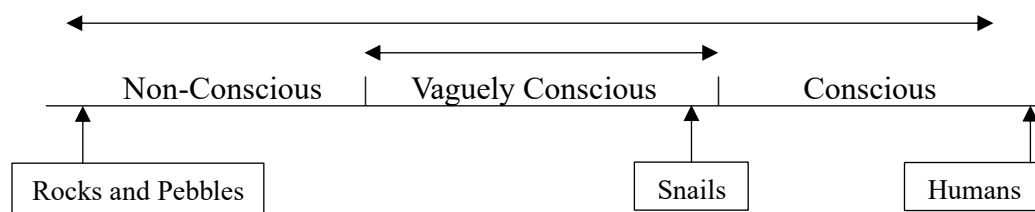
conscious.⁷⁵ For this third category of entities, it is neither determinately true nor determinately false that there is something it is like to be those entities in question. We can think of such a category of entities as occupying an indeterminate ‘grey zone’ of consciousness (Schwitzgebel, 2023: 3415-3416). Now, it could be that this grey zone is defined by entities that are relatively low on the scale of neural and cognitive complexity, or conversely, by those that are relatively high on this scale (but yet still more cognitively simple than us). Similarly, it could be the case that this grey area is narrow, such that only a few species fall into it, or that it is broad, such that the majority of living things fall within it. Indeed, the depth and range of any of our three categories could be infinitesimally small or large, and how we should define them (and indeed, what entities will fall into which) will be entirely dependent upon which theory of consciousness turns out to be true, and how that truth manifests in the world. For the readers benefit, however, we have chosen our token entities for each category to include pebbles as determinately non-conscious, snails as neither determinately conscious nor non-conscious, and humans as determinately conscious.

However one wishes to classify the depth and range of our grey zone category, we should say that no sharp line divides the conscious organisms from the non-conscious ones (ibid: 3418). Similarly, we should say that no sharp line divides the conscious category from vaguely conscious category, and from that vaguely conscious category to the non-conscious category. Instead, and per Schwitzgebel, there is sort of ‘gradual fading’ or gradation all the way down, with all entities and things falling into one of the three listed categories. Perhaps humans are at the top of the determinately conscious category, snails are somewhere near the top of the vaguely conscious category, and pebbles are towards the bottom of the determinately non-

⁷⁵ Crucially, if consciousness is vague then it would be vague where such boundaries are and how many entities fall within each classification. However, we should still theoretically posit such boundaries and such relevant theoretical entities, so as to account for these necessary classifications.

conscious category. Perhaps it is neither determinately true nor false that there is something it is like to be a snail, but whether or not there is something it is like to be a snail is less vague than, for example, a protozoan. In this manner, its consciousness is more vague than ours, but less vague than the protozoan's.⁷⁶

We might highlight these clarifications onto our trifurcation as follows:



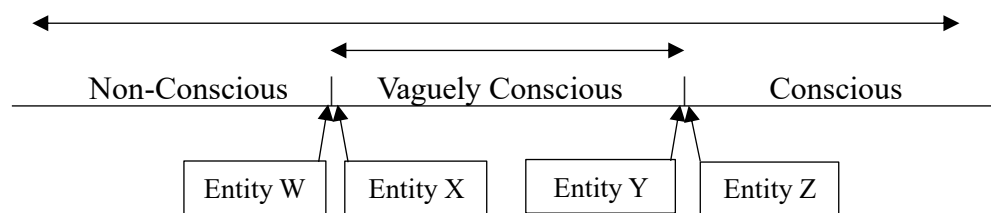
With these clarifications in place – and with particular emphasis on the fact that the depth and range of any of our three categories could infinitesimally small or large – we can now focus our attention to theoretical borderline cases within our trifurcation. Let us take some theoretical ‘entity y’, for which it is neither determinately true nor false that there is something it is like to be it, but which sits further towards the determinately conscious category than all other entities who are vaguely conscious.⁷⁷ Let us compare this with a theoretical ‘entity z’, for which it is determinately true that there is something it is like to be, but which sits lower on the spectrum of ‘determinately conscious’ than all other entities in its category. We can say that entity y and entity z sit next to each other on the spectrum yet fall into separate classifications: the latter is determinately conscious, the former is vaguely conscious.

⁷⁶ We should assume that the ‘gradual fading’ extends into the vague category simply to account for general intuitions regarding consciousness. One might think that both a protozoan and a snail would fall into the vaguely conscious category. But, knowing what we know about either entity’s behaviour and constitution, intuitively, and per Schwitzgebel’s reasoning, the snail is closer to the determinately conscious category than the protozoan.

⁷⁷ The physicalist may reply that the boundaries can also be vague, such that there is not a point on the spectrum which marks the furthest point. I shall consider this objection shortly.

Similarly, let us take some theoretical ‘entity w’ that occupies the highest segment of the determinately non-conscious category. This is an entity for which it is determinately false that there is something it is like to be, but it sits higher on the spectrum than all other entities in its category. Let us compare this to some theoretical ‘entity x’ that occupies the lowest segment of the vaguely conscious category. This is an entity for which it is neither determinately true nor false that there is something it is like to be, and which sits lower on the spectrum of the vaguely conscious category than all other entities in that category. Like those entities above, we can say that entity w and entity x sit next to each other on the spectrum yet fall into separate classifications: the latter is vaguely conscious, and the former is determinately non-conscious.

Let us highlight these theoretical entities on our trifurcation as follows:



We can now ask the physicalist to reflect on what these borderline cases would mean for their theory. In particular, what is it about entity z that places it in a distinct category to entity y? Similarly, what is it about entity x that distinguishes it from entity w? Since our spectrum of classification forms a gradual fading such that no sharp line divides any category from the next, then in practice, there should be no sharp line in physical constitution that separates x from w, and z from y, respectively. What kinds or types of entity w-z are is unimportant to the discussion here. What is more important is the fact that whatever entities they are or turn out to be, since there is a ‘gradual fading’ within our spectrum of classification, the difference between w and x, and between y and z, should be, in principle, one of a fundamentally small degree. As a result of this, it might be argued that the physicalist must concede that the difference between a

determinately non-conscious entity and a vaguely conscious entity, alongside the difference between a vaguely conscious entity and a determinately conscious entity, is one of the smallest fundamental degree.

The physicalist must then ask himself what it is about those fundamental differences that is relevant to those entities' different classification of category. Is it the addition of one extra fundamental part? Is it a marginally different configuration of the same fundamental parts? However the physicalist responds, what becomes apparent is that there must be something about those fundamental parts relevant to the change in categorisation. Furthermore, since these categorisations consider levels of consciousness, we can say that for the physicalist, there must be something about those individual fundamental parts relevant to the instantiation of consciousness, and this is a distinctly panpsychist move, and at odds with our physicalist's account of fundamentality. For example, if entity w can become vaguely conscious in a manner akin to entity x (or if entity y can become determinately conscious in a manner akin to entity z) upon the addition of one extra fundamental part or some reconfiguration of existing parts, then there must be something about that additional fundamental part relevant to instantiating consciousness, or something about all the existing fundamental parts relevant to instantiating consciousness when appropriately combined. In this manner, if the physicalist is to reconcile their thesis with the vagueness of consciousness, they must attribute the property of consciousness – or properties relevant to instantiating it – at the fundamental level. However, in doing so, the physicalist will unintentionally find themselves endorsing something genuinely similar to Russellian panpsychism. Unlike in Schwitzgebel's objection, this should indeed worry the physicalist.

The physicalist who still wishes to endorse the vagueness of consciousness (and also wants to avoid the above conclusion) is more likely to propose that the gradual fading that exists within our trifurcation means that our reference to any categories within it is mistaken. Along these lines, they might argue that there is vagueness in constitution, such that it is unfair to ask for some positive characterisation between those at the top of the non-conscious category and the bottom of the vaguely conscious category, and between those at the top of the vaguely conscious category, and the bottom of the determinately conscious category. Instead, there might be only vague physical differences between those borderline cases, such that reference to some absolute difference becomes impossible.

In response, the Russellian panpsychist may reemphasise that the borderline entities of concern to this section are entirely *theoretical*, such that even if there are only vague differences between borderline cases in practice, there should be some determinable difference in principle. Without this difference in principle, the Russellian panpsychist may propose that this physicalist rebuttal faces a particularly strong Sorites paradox. Traditionally, Sorites paradoxes are formulated around a vague concept or property, for example ‘heap,’ and illustrate the impossibility of defining any precise moment of transition from one classification to another. For example, one must acknowledge that a single grain of wheat does not make a heap; nor do two grains, three grains, four grains, and so on. However, if n grains of wheat do not constitute a heap, then $n + 1$ grains do not either. Accordingly, no matter how many grains of wheat one adds, they never constitute a heap (Horgan 1994: 159-188; Soames 1999). Conversely, let us consider a genuine heap of wheat: removing one grain does not mean the structure is no longer a heap; nor does removing two, three, four, and so on. However, if removing n grains of wheat does not destroy the heap, then removing $n + 1$ grains also never destroys the heap; even when the heap is reduced to a single grain of wheat, it must still be a heap.

Sorites paradoxes appear to validly yield a falsehood: a single grain of wheat is obviously not a heap of wheat. While it is beyond the scope of this thesis to rule out all proposed solutions to the paradox, we can see a clear sense in which the paradox appears to show that ‘vagueness,’ strictly speaking, is a logically incoherent notion. If this reasoning is correct, vagueness as features of objects or properties is impossible, and so properties like ‘bald,’ and ‘heap’ don’t really exist in the world (Horgan, 2010: 67). Instead, most philosophers choose to address the paradox semantically, such that vague concepts and properties exist solely in our language, whereby they provide a sort of useful fiction (Heller, 1998: 111). Indeed, while properties such as ‘heap’ and ‘bald’ may be incoherent, we can still see how they provide a useful way of navigating reality that is viable, legitimate, and indispensable (Horgan, 2010: 67). Accordingly, while Sorites paradoxes remain interesting, they don’t really matter for practical purposes regarding our traditionally vague notions; concepts like ‘bald,’ or ‘heap,’ can simply apply to objects in degree or as context dependent predicates, rather than strictly true or false terms. Such linguistically vague terms remain very useful for negotiating a (non-vague) reality that genuinely is there, and therefore, their subtle incoherence does not typically cause an issue.

With this in mind, let us now apply the paradox to consciousness. Assuming that consciousness is vague, it is then vague as to what physical difference or change is sufficient to change a nonconscious entity to a vaguely conscious entity, and a vaguely conscious entity to a conscious one. However, and as illustrated by the paradox, there can also be no physical difference sufficient to change a conscious entity to vaguely conscious entity, nor from a vaguely conscious entity to nonconscious one. Since the physicalist only knows with certainty that they are conscious – as opposed to knowing that other things are not conscious – it is this latter formulation of the paradox that is of worry to the physicalist. Since we are determinately

conscious, then the removal or change of one relevant fundamental part does not change it to a vaguely conscious entity (or a nonconscious entity), nor does the removal of two, three, or four. However, if the removal or change of n relevant fundamental parts does not make a vaguely conscious entity (or a nonconscious entity), then the removal of $n + 1$ fundamental parts does not either. In fact, while such removals or changes may produce entities which are less and less determinately conscious – and so further towards the vaguely conscious category – we can never get to vaguely conscious or nonconscious entities within our spectrum of classification.

Continuing to assume that consciousness is vague, one might suggest that consciousness must then also be an incoherent notion, such that the property ‘consciousness,’ strictly speaking, does not exist; its instantiation seems to imply the contradictions revealed by the Sorites-style reasoning above. Now, while this conclusion may be contentious, it is certainly a potential objection that the physicalist who endorses the vagueness of consciousness might run into. Furthermore, while this reasoning did not pose any real issue for our traditionally vague notions, such as ‘heap,’ or ‘bald,’ if such analysis is valid, it does pose a big issue for the physicalist. First and foremost, consciousness definitely *does* exist; it is not just a useful fiction. Accordingly, the Sorites paradox might force the physicalist to deny the very phenomena they sought to explain and thus violate the tenets of both type-b and -c physicalism of interest to this thesis. Furthermore, the Consciousness Constraint within the Mind-Body Problem entails the requirement that any adequate theory of reality must entail that consciousness is explained. In this manner, if the physicalist endorses the vagueness of consciousness, they might fail to satisfy the Consciousness Constraint and so fail to provide any solution to the Mind-Body Problem. Their only other option is to conclude that the removal of $n + 1$ fundamental parts never changes a conscious entity to a nonconscious one, and thus conclude that everything is,

to varying degrees of complexity, conscious. As before, this is a distinctly panpsychist conclusion, and one in which any physicalist should seek to deny.

So as to avoid running into these sorts of potential issues, I propose that the physicalist should simply endorse that consciousness is always precise. We have already considered the good intuitive reasons we have for this conclusion. As we saw, the physicalist cannot provide a positive characterisation of any borderline instance of consciousness, nor any potential ‘sharpenings’ to the predicate ‘is conscious.’ With this in mind, there appeared to be little reason to suppose that consciousness is vague from the offset. We then assessed the potential consequences that physicalism might run into should they continue at this point to deny its precision. In particular, we proposed that either the physicalist must give a positive characterisation for the difference in classification between entities in the nonconscious/vaguely conscious/conscious categories, or they must endorse vagueness in constitution, which forces a Sorites paradox. While the former option potentially threatens a reduction into panpsychism, the latter option could be used to threaten physicalism’s ability to effectively respond to the Mind-Body Problem. As noted, this argument was a ‘backup argument,’ and so if the physicalist simply chooses to endorse that consciousness is precise, we need not reach this conclusion. All other things considered, let us simply grant that the physicalist should endorse the precision of consciousness.

5.3.2 Evolution, Saltation, and Biopsychism

Maintaining the above assumption, we can now ask the physicalist, “When did the light turn on?” or “When did phenomenal experience ‘pop up’ within our evolutionary trajectory?”.⁷⁸

⁷⁸ I shall avoid the question of “Why?”, as this might be answered away in terms of randomness and evolutionary benefit.

Continuing under the assumption that consciousness is precise, there are two clear options available. The first is to propose that phenomenal experience is a result of saltation (Schwitzgebel, 2023: 3418). This option endorses the clear distinction between conscious and non-conscious things and its proponents should, in theory, be able to point to a time within our evolutionary trajectory where consciousness ‘popped up’ as a sudden and large mutational change. The second option is to propose that phenomenal experience was always there, built into the fundamental parts of the universe (ibid). This option endorses panpsychism, and its proponents should expect to see a clear evolution of conscious things into further conscious things. Now, clearly the second option is again unpalatable for the physicalist, and so we shall turn to the first option instead.

Pursuing the first option, the physicalist must look for a ‘line in the sand’, where the determinately nonconscious mutates to become the determinately conscious within our evolutionary trajectory. For the type-c physicalist, this line should illustrate some relevant physical difference between conscious and nonconscious entities that should in principle be relevant to explaining away the production of consciousness. For the type-b physicalist, while this line may not distinguish a genuine physical difference that will help explain away the Hard Problem, it should at least point to a relevant physical difference or correlation indicative of the primitive identity relation between physical and phenomenal facts. Crucially, both variants of physicalism are searching here for a *non-arbitrary* point for the origin of consciousness within evolution. As noted by Clifford and James, such a point needs to illustrate a sufficient physical difference between nonconscious things and conscious things, so as to account for the occurrence of consciousness in the first place (Clifford, 1874: 61; James, 1890: 148).

Now, if consciousness is truly precise, then it would likely be the case that one cell – or even one particle – would make all the difference.⁷⁹ However, if it is truly the case that one single cell makes all the difference, and that this cell is also truly non-arbitrary, then physicalism might again reduce into something similar to panpsychism (for the same reasons as those presented to the physicalist who attempted to find a relevant physical distinction between borderline entities in our spectrum regarding vague consciousness). So as to avoid this conclusion, the physicalist should, therefore, look to broader physical differences between potentially conscious and non-conscious entities, such as some distinguishing physical structure, within which no specific single cell makes all the difference in practice, but where the structure needs each and every cell so as to function in principle. Within this sort of non-arbitrary structure, the physicalist can say that a single, non-specific cell makes all the difference, and so retain precision without necessarily adding any further metaphysical commitment.

Rather than now having to isolate some specific single cell that is sufficient to illustrate the origin of consciousness, the physicalist needs only to find some sufficient and distinguishing physical structure. Let us now actively look for such. Now, we know with greater certainty what things are definitely conscious than we know which things are definitely not. In order to find this line in the sand then, we should first start from those things that we *know* are conscious and consider whether we can trace evidence for it further down our evolutionary origin, and whether we can find sufficient physical difference between those things that are conscious and

⁷⁹ The physicalist who endorses Integrated Information Theory (IIT) might chose to endorse this claim (Tononi, 2012; Koch & Tononi, 2015; Koch, 2014). IIT posits that consciousness is an intrinsic fundamental property of reality that can be found in small gradable amounts in simple physical systems, and have also proposed a mechanism to articulate which systems are conscious and which are not. However, few type-b and -c physicalists endorse this theory due to its similarities to variants of panpsychism, and it has also been generally shunned as pseudoscience within academia (Fleming et al, 2015; Cerullo, 2015). As such, I have abstained from commentary on this potential objection.

those things we think are not. With this in mind, let us look to distinctions between sub-species, and also species as a whole, so as to assess whether there is any clear distinction in constitution that could delineate conscious things from nonconscious ones. If we should find such a non-arbitrary distinction, then we can begin to assess what it is about that constitution that supports the delineation between conscious and nonconscious things. Crucially, this non-arbitrary distinction should exist in principle for both type-b and type-c physicalism, as both theories endorse that there exist conscious things, and nonconscious things.

Now, it is clear to me that I am conscious: I am having an internal experience of my thoughts, the keys under my fingers, and the sound of my tapping them. Perhaps then, the first candidate for consciousness originating should be with us: *homo sapiens*. The first potential line in the sand would, therefore, be between *homo sapiens*, and everything else. Since consciousness and the brain are clearly intrinsically linked, we should look to whether there is any clear distinction between the precise structure of our brains and every other entity, that could signify a nonarbitrary delineation between conscious and nonconscious things. However, while *homo sapiens* are certainly a clear candidate for what *is* conscious, claiming that it arose only within them would be too hasty, and against the intuition of most. With shared reasoning to James and Clifford, there simply does not appear to be any non-arbitrary physical difference between *homo sapiens* and their closest evolutionary relatives – such as the *homo neanderthalensis*, or *homo erectus* – where consciousness should intuitively originate in. For example, *homo neanderthalensis* had comparable brain sizes to modern humans, and differed structurally only inasmuch as they were more elongated and less globular than *homo sapiens* (Gunz et al., 2010). Since the only clear and relevant physical difference exists in shape, the physicalist who wishes to maintain that only *homo sapiens* are conscious must endorse that consciousness finds its origins in brain shape: a truly arbitrary distinction.

The physicalist who does not want to endorse such an arbitrary distinction might instead attribute consciousness to all our closest evolutionary relatives (such as those listed above) and place a line in the sand between them and their more distant relatives, such as non-human primates. Following this reasoning, it could be that those of the genus *homo* are conscious, but that non-human primates (and everything else) are not. In order to justify this distinction, we should look to those precise structure and features of non-human primates that might represent a relevant non-arbitrary difference to us with regards to the production of consciousness. For example, while gorillas and orangutans are at least as large as *homo sapiens*, their brains are at least one third of the size of the human brain (Herculano-Houzel & Kaas, 2011). In this manner, the physicalist might say that the genus *homo* is special for its exceptionally large brain to body size ratio, and that this is the non-arbitrary distinction between our consciousness, and our non-conscious distant relatives. However, Herculano-Houzel and Kaas have suggested that rather than the human brain being special in size, it is the great apes who have evolved bodies that are unusually large. In fact, humans possess a brain that matches what is expected for a primate of their size, such that it is linearly scaled up primate brain in its number of cells (ibid). In this manner, the size of the human brain, or proportional size of it to our bodies is an arbitrary distinction from our relatives when concerning consciousness production: our brain is no more special in its size or cellular constitution than that of most other primates.

The physicalist might instead then point to some alleged evidence for a lack of consciousness in such non-human primates, before then trying to establish some physical evidence for *why* they are not conscious, and we are. However, recent studies should serve to negate this option entirely. For example, in a recent study by Günter Ehret and Raymond Romand (2022), macaque monkeys were shown to possess ERP visual awareness negativities with similar

properties to humans, in such a fashion that they concluded that they are subjectively aware and conscious about perceived stimuli (Ehret & Romand, 2022). Crucially, this consciousness is suggested to be representative of higher-order consciousness, as awareness requires *thought about* those stimuli, as opposed to simply *experience of* them. In this manner, the physicalist cannot point to a lack of evidence for even complex consciousness in non-human primates, let alone a lack of evidence for consciousness, simpliciter.

The physicalist might instead attempt to place the line in the sand at the division between primates and other mammals. However, again, such a division has little basis on the grounds of relevant physical differences. Humans, let alone primates, do not rank even close to first in relative brain size (when expressed as a percentage of body mass) within the mammalian kingdom, in absolute size of the cerebral cortex, or in gyrification (Hofman, 1985; Herculano-Houzel, 2013: 128). Furthermore, not only is it a relatively uncontentious household assumption that many (if not all) mammals are conscious, but there is also ample evidence to support such assumptions. For example, investigations targeting self-recognition, self-awareness, and metacognition have shown that dolphins can recognise themselves in mirrors, as well as to monitor their own knowledge in perceptual categorisation tasks (Marino, Reiss, Gallup, 1994; Marten & Psarakos, 1994; 1995a; 1995b). Like primates then, not only should it be uncontentious that (at least some) mammals possess consciousness, but it is also uncontentious that at least some possess complex forms of consciousness. In this manner, the physicalist cannot point to any non-arbitrary point within our evolution away from our mammalian relatives relevant to the origin of our consciousness.

I'm sure the reader is starting to get the picture: at least within those animals similar to us in their evolutionary trajectory, there is no non-arbitrary point in our evolutionary development

away from them that the physicalist can point to that should signify the origin of consciousness, nor sufficient reason to suppose that those similar animals are not conscious in the first place.⁸⁰ Continuing under the assumption that the reader sees no reason to assert our mammalian relatives are not conscious, nor sees any sufficient physical grounds to defend such, let us skip out much of the further spectrum of relatives left to discuss, and instead attempt to draw our line in the sand – of those things that are conscious and not – between those things that have brains and those that do not.

Many past attempts to define a ‘brain’ have suffered from a human bias, for example, requiring a particularly convoluted structure that might render obviously brained genus’ such as rodents, birds, fish, amphibians, reptiles, and even some humans with certain neuroblast migratory disorders as brainless (Sarnat & Netsky, 2002: 241). Others, have focussed their definitions more philosophically on the ability to think, reason and solve problems, which might also easily exclude many obviously brained creatures (ibid). For this reason, let us forgo a strict definition of a brain, and focus instead on how it is that the things we call ‘brains’ evolved.

Now, it is clear that we have evolved from a long line of ancestors that were evermore less similar to ourselves (Kaas, 2013: 34). For example, our early mammalian ancestors had on the order of 15-20 cortical areas, current macaques have as many as 129, and modern-day humans have anywhere in the region of 150-200 (ibid: 36-42). Clearly then, brains have slowly and incrementally evolved to be more complex, and more specialised. The issue with ascribing consciousness to only things that have brains is, therefore, that the evolution of brains was a developmental process: at some point in our ancestor’s evolution, they evolved neurons; closer

⁸⁰ While my reasoning here may be uncontentious to the reader, it is worth noting that near to half of all the countries in the world do not recognise animal sentience, simpliciter (See <https://api.worldanimalprotection.org/> for more details). With this in mind, while most may find my claims uncontentious, they needed to be explored, nonetheless.

ancestors then evolved simple nervous systems; and, finally, closer yet ancestors evolved more evolutionary advanced brains. Not only then is there no clear point in our ancestor's evolution where we can point to and say, "that's where we evolved a brain," there is also no clear point where we can say, "that's where we evolved a brain *and* consciousness." The question that then stands for the physicalist is, therefore, this: at what point within this developmental process was our structure sufficiently complex so as to produce consciousness? Since the brain's evolution was an incremental, gradual development, there was no sufficiently large 'jump' from one ancestor to the next with regards to the development of their brains, and certainly not one that could intuitively account for the introduction of consciousness.

Like the slow and incremental evolution of brains, the evolution from our most distant ancestors into modern day humans has also been slow and incremental. *Homo sapiens*, and all those organisms and genera mentioned prior, arose as a fortuitous and contingent outcome of an unsurmountable number of linked events. Natural selection is a principle of local adaption, where living things 'struggle' among and against other organisms for reproductive success, leading to an improved fit of populations within changing environments (Gould, 1994: 85). For this reason, all evolutionary changes occurred very slowly. For example, in our lobe-finned fish ancestors, those with stronger fin bones possessed an evolutionary advantage that was slowly and incrementally passed on to their offspring, resulting in their eventual development of a strong central axis capable of bearing weight out of water. Similarly, and in our more recent primate ancestry within the African savannas, those with more vertical postures were more successful at harvesting resources, and so surviving and reproducing. As such, postures incrementally became more vertical, culminating in our more recent ancestors eventually standing entirely upright (ibid: 86). We can say then that our entire physical development, from

our most distant to our most recent ancestors – like the development of brains – has been slow and incremental at every stage.

Those physicalists unwilling to endorse the claim that some single cell could make all the difference (but who still wish to find a line in the sand between conscious and non-conscious things) will eventually, therefore, be led to the conclusion that perhaps all living things are conscious, and all non-living things are not. Since the physicalist appears to fail at finding a non-arbitrary point within our evolutionary development that would signify the introduction of consciousness, it would follow that our more mechanically simple ancestors also enjoyed phenomenal experience. This trend can be continued down our entire evolutionary ancestry, all the way back to the first living things; just like the evolution of brains, there was never a sufficiently large physical change between one creature to the next that would signify the introduction of consciousness, and so consciousness must have been there at the beginning of the evolutionary picture, in what has now become known as LUCA (the last universal common ancestor).

Information on how and where LUCA lived is lacking. The most popular evidence suggests an autotrophic origin of life involving the Wood–Ljungdahl pathway in a hydrothermal setting (Weiss et al, 2016: 1). However, we need not burden ourselves here with an inquiry into the vast literature surrounding the topic. What is of sole importance to this analysis, is that whatever LUCA was, and wherever LUCA lived, it is the most distant ancestor that all living things share, and so intuitively marks the beginning of life – as we know it – on earth, and the beginning of the slow, incremental development that we resulted from. Failing to find any non-arbitrary point in our evolution that would signify the introduction of consciousness, the physicalist is, therefore, led to the conclusion that LUCA must have also had consciousness, so

as to pass on that property throughout evolution into more complex and sophisticated forms. Crucially, for the physicalist to say that LUCA must have had consciousness is not to say that it must have had a rich and complex inner life. Instead, to posit consciousness at the beginning of the evolutionary picture is only to posit that there was simply something it was like to be LUCA.

To say that consciousness was present in LUCA is to endorse ‘biopsychism,’ or the enactive conception of life (Thompson, 2007; Di Paolo, 2018). Coined by Ernst Haeckel at the end of the nineteenth century, ‘biopsychism’ initially maintained that *all and only* living organisms are conscious, such that the same concepts required for explaining the phenomena of life are required for explaining consciousness (Haeckel, 1892; Thompson, 2007; Di Paolo, Buhrmann and Barandiaran, 2017; Di Paolo, Cuffari and De Jaegher, 2018). Contemporary biopsychists, however, have reformulated this thesis to make a weaker claim. They wish to permit the possibility of, for example, Strong Artificial Intelligence, or non-organismic sentience, and so to avoid the initial identity relation proposed by Haeckel; contemporary biopsychism simply maintains that *at least* all living things are conscious (Thompson, 2022: 231). Because of this weaker claim, biopsychism opens itself to be grounded by further metaphysical theories, such as different variants of dualism, or monism. For example, the biopsychist could chose to ground their theory in panpsychism, such that the combination of fundamental conscious parts occurs in at least all living things. Or they could ground it in substance dualism, such that at least all living things have two substances, one mental, and one physical.

The physicalist who wishes to suggest life as the non-arbitrary point for which consciousness begins, therefore, becomes a biopsychist. They would maintain that there is a genuine physical difference between most living and non-living things and that this difference represents the

non-arbitrary point at which consciousness begins. Crucially, this is the final avenue they can take in their search for a non-arbitrary difference while maintaining physicalist standards; if they wish to suggest the consciousness of anything ‘lower’ than life, they must suppose that inanimate things are conscious, and such a concession would leave no demarcate difference between their own assertions, and the panpsychist implications they hope to deter. As we shall see in the next subsection, however, the type-b and -c biopsychist physicalist runs into some grave issues, that not only worsen their hard problem, but also introduce numerous further problems that advance the claim of the Russellian panpsychist.

5.3.3 The Issue with Biopsychist Physicalism

From the offset, the physicalist might already be weary of their new biopsychist position. In particular, the biopsychist position appears to be much less restrictive with regards to its ontological claims about consciousness than the physicalist would typically like to make. For example, if we assume that consciousness is evolutionarily beneficial to any possessor, then it is not likely something that would have ever been de-evolved, and so if LUCA was conscious, then everything that evolved from LUCA is probably conscious too. This would mean that for the biopsychist physicalist, protozoans, plants, insects, and so on, should all be classified as conscious things. Now, the physicalist might not intuitively wish to endorse this claim. However, in light of the prior analysis in this regard, it is not clear what other options he has available at his current disposal: for the physicalist who wishes to endorse the claim that consciousness is precise, life appears to be the first non-arbitrary place one can point to for the origin of consciousness, but also perhaps the last non-arbitrary place they can endorse before their doctrine dissolves into something similar to panpsychism.⁸¹

⁸¹ One might be immediately concerned as to whether life truly is a non-arbitrary point. Unless moving an electron marks the boundary between life and non-life, life itself might be a vague concept that cannot represent a non-arbitrary origin for consciousness. This worry will be returned to later. For the moment, it will suffice to

Unfortunately for the physicalist, this is not the only issue that their new biopsychist position runs into. First, if consciousness is now synonymous with life, the biopsychist physicalist still needs to give an account for *why* those most basic living things are conscious. As noted previously, the first required step was to find the point at which consciousness begins accompanying the physical parts and processes which allegedly necessitate or identify with consciousness; the second step was to explain what it was about those physical parts and processes that were relevant to the production of consciousness. Now, this step can only be asked of the type-c physicalist. Type-b physicalists do not endorse that the epistemic gap encapsulated in the Hard Problem would be explained away; they believe the identity relation between consciousness and physical parts and processes is primitive. However, type-c physicalists argue that the epistemic issue *will* come to be explained away as we uncover more about the mysteries of consciousness and its relation to the physical. In this regard, they must explain what it is about living things that is relevant to the existence of their consciousness.

The issue for the type-c physicalist in this regard, is its particular implications to their Hard Problem. Now, while their claim that the Hard Problem is a challenge yet to be explained away might have some intuitive merit when considered against the vast complexity of the physical brain and body, the processes in, for example, a single celled organism, are comparatively simple, well-explored and well-understood. In this regard, if every living thing is conscious – including those most simple living things – then the answer to the Hard Problem should be accessible with a knowledge of the parts and processes of those most simple things. The type-c physicalist can, therefore, no longer appeal to the abstract complexity of brains and bodies,

say that life is the *last* point (arbitrary or not) that the physicalist can point without their theory reducing into something similar to panpsychism.

nor the vast physical data yet to be explored as the reason for the existence of their Hard Problem; their Hard Problem now requires a simpler answer that appeals to the constitution of simple living things.

We have a relatively robust knowledge of the relevant physical constituents of the simplest living things. Take, for example, the single-celled protozoan paramecium. We understand the qualities of the molecules which compose them, and how their particular biological arrangement gives rise to behaviour, without any of the additional complexity observed in multicellular organisms (Delafield-Butt, 2007: 238). For example, the smallest of these elements are free ions and atoms, which form the cell's internal cytoplasmic solutions and are accompanied with free amino acids and nucleotides in the cytoplasm. Dozens to hundreds of amino acids then make up the proteins that in turn make up mitochondria, cytoskeletal structures, the nuclear envelope, cytoskeletal structures, plasma membrane, and so on (ibid). The particular arrangement of these variables and their function is well documented, and more importantly, comprehensible to the human mind: there is no epistemic gap between our knowledge of paramecia constitution and of paramecia action and behaviour.

The question for the type-c physicalist is then this: how can knowledge of these simple entities explain away the Hard Problem? If – as the type-c physicalist has now conceded – these entities are conscious, then we should be able to see something about their constitution which explains away their consciousness. However, at this level of enquiry, and concerning entities of which our knowledge is relatively complete, the Hard Problem still remains; there is no clear reason for *why* these entities are conscious. In this regard, the necessary variables relevant to explaining away consciousness – à la type-c – have become more restricted and simpler, and yet the gap in our explanation has got larger. Indeed, since the type-c physicalist could

previously appeal to the great complexity of the human brain and body, the claim that such complexities may contain an unexpected solution to the Hard Problem was perhaps more plausible. However, if the Hard Problem persists in simple living things – of which we have a robust understanding of – it is far less likely for their simple constitution to yield a surprising explanation to the problem; it appears that their Hard Problem has now become insurmountably harder. Accordingly, the claim that the Hard Problem could ever come to be explained away appears to be at best, unlikely, and at worst, entirely redundant. In this manner, the type-c biopsychist physicalist seems to have forced the undermining of their entire thesis.

Failing to explain away the Hard Problem at the level of simple living things – and since they cannot turn to more complex living things in the hope they will explain away the Hard Problem – the only clear course of action available to the type-c physicalist is to state that there must be something about the nature of the parts which constitute those living things that disposes or is relevant to instantiating consciousness when appropriately combined. However, and with regards to the simplest living things, this requires that non-living parts are now relevant to instantiating consciousness. As before, this concession would leave no demarcate difference between type-c physicalism and the Russellian monist implications (be they panpsychist or panprotopsyichist) it hopes to deter. In this manner, it might simply be easier for the type-c physicalist to abandon their thesis and endorse Russellian panpsychism so as to remove the Hard Problem entirely.

Interesting, the relationship between biopsychism and the Hard Problem has not received any sustained discussion within the academic literature. However, it is clearly an issue that contemporary biopsychism might run into independently of physicalism: if all life is conscious, then there must be something about life relevant to consciousness. But, since we understand

the constitution of the simplest living things – in particular, how that constitution is relevant to their *living* – there appears to be little to no gap left in our explanation. In this manner, biopsychism could be burdened with the same severity of Hard Problem as our type-c biopsychist physicalist was. Now, perhaps the biopsychist can claim a primitive identity relation between living things and conscious things, and so claim that the epistemic gap will never be explained away. Or they could ground their theory in something similar to panpsychism, and in doing so, build simple consciousness in *below* the level of living things, in perhaps, a non-combined sense, and posit mental combination to begin occurring at the point of life. However, such is beyond the particular purview of this chapter, and, as noted above, is of no help to the type-c physicalist here.

The second issue that the biopsychist physicalist will run into concerns both the type-b and type-c variants. In positing that all living things are now conscious, the type-b and -c physicalist will run into a similar ‘Combination Problem’ to Russellian monism. The human body is understood to contain approximately 37.2 trillion human cells (Bianconi et al, 2013). Crucially, each of these cells satisfy our definition of a living thing, and so, by our physicalist’s concession, could now be viewed as individually conscious entities. However, if each of the 37.2 trillion human cells in our bodies are conscious, then how is that they combine into a singular unified experience at our macro-level? As proposed to the Russellian panpsychist, experiences are not intuitively of the right kind to form aggregates. To modify James’ claim, “Take a hundred [living conscious things], shuffle them and pack them as close together as you can (whatever that may mean); still each remains the same [living conscious thing] it always was, shut in its own skin, windowless, ignorant of what the other [living conscious things]’ feelings are and mean.” (James, 1890: 180, [my amendments]).” Or, in the words of Lucretius “Still by their meeting and union they will make nothing but crowd and mob of living things,

even as, as you may know, men, herds of cattle and wild beasts could not beget anything by coming together with one another” (Lucretius, 1910: 73). In the same manner as Russellian panpsychism then, the physicalist appears to inherit a Combination Problem of similar form: humans are now composed of ‘little’ conscious things, which combine to make a ‘larger’ conscious thing, and this should pose at least as much of a metaphysical problem for the physicalist as it did for the Russellian panpsychist.⁸²

This issue presents itself to both type-b and type-c forms of biopsychist physicalism. The type-b physicalist accepted that the Hard Problem is an epistemic gap that will not come to be resolved but denied that the problem is an ontological one in nature. Instead, they identified conscious states with physical states. While they maintained that this identity relation is primitive, such that it cannot be explained, since all living things are now conscious, they must now explain how the conscious states at, for example, the (living) cellular level, explain the conscious states of the (living) whole. The matter of how these conscious cells come to compose macro-consciousness is no longer just a matter of primitive identity; it necessitates a combination mechanism that explains how multiple conscious things create a further conscious thing. The type-c physicalist, on the other hand, proposed that the Hard Problem will come to be explained away. While this problem has been shown to have been worsened by their new biopsychist position, they are now also required to explain how the multitude of conscious cells in our body combine to form a singular and unified consciousness.

Now, this new Combination Problem does not in-and-of-itself possess more dialectical force than the similar argument proposed against Russellian panpsychism. However, it does

⁸² Now, the physicalist *could* opt to resist that macro consciousness is formed of micro consciousness in such a way that although the parts are conscious, the consciousness of those parts does not constitute the consciousness of the whole (just like emergentist panpsychism). However, this would amount to a strong emergentist form of physicalism, and I have opted to remove all forms of strong emergence from my analysis.

introduce a further problem that the physicalist must now resolve. In this manner, not only does the Hard Problem still linger, they have now also inherited the most dangerous argument against Russellian panpsychism. I began this chapter by emphasising the theoretical ‘level playing field’ between Russellian panpsychism and physicalism. Both had respective and unresolved problems – the Combination Problem, and the Hard Problem, respectively – and I proposed that one or the other theory must show how their respective problem is not a problem at all, or that the other’s problem is *more* of a problem than their own. It is clear that through adopting biopsychism, physicalism no longer enjoys this level playing field; the physicalist position now possesses the Combination Problem *and* the Hard Problem, and the Russellian panpsychist only possesses the former. Therefore, while the Combination Problem for physicalism is not damning in isolation, it should certainly force a shift of intuition regarding the individual merits of either theory.

Unlike the biopsychist Hard Problem, the biopsychist Combination Problem is acknowledged within the extant literature; it is understood to be one the big challenges facing the theory (regardless of whether or not biopsychism is paired with type-b or -c physicalism) (Ginsburg and Jablonka, 2020: 21; Thompson, 2022: 249). For example, commentators have acknowledged that if life suffices for consciousness, then the question of what happens to individual cellular consciousness in the case of multicellular organisms, cell colonies, or quorum sensing in bacterial populations must be answered. For these commentators, the question is whether the individual constituent cells retain their consciousness, or whether this consciousness is subsumed and integrated into the multicellular individual (ibid: 249-250).

Arthur Reber and František Baluška (2021) have begun to address this problem for biopsychism.⁸³ Reber and Baluška propose firstly that the individual cells in a multicellular organism retain their own consciousness while contributing to the larger consciousness of the organism. This model entails that proposed implicit learning and memory independently contribute to conscious awareness and depend on evolutionary older cortical systems and pathways. Similarly, individual cell consciousness takes place independently of, but contributes to, the consciousness of the multicellular organism. In this potential explanation, the macro-organism does not experience the conscious states of its individual cells. Instead, its consciousness results from integrating them into a ‘larger Gestalt’ (ibid: 152).

Reber and Baluška’s model is reminiscent of Leibniz’s (1989) idea that when listening to the sound of a stormy sea, all we hear is a confused integrated roar of all the waves. When hearing this roar, we hear each individual wave, but we cannot recognise each individual wave in sound, we only know them through their integration into the overall sound. Now, while this does not itself constitute a resolution to the Combination Problem – for it simply focuses on *what* happens, rather than *how* it happens – it signifies a start of a potential resolution, nonetheless.

Since biopsychists do themselves acknowledge the Combination Problem, it certainly a problem that both type-b and -c biopsychist physicalism must also. Now, both variants could endorse and develop the sorts of potential responses initiated by biopsychists, or they could attempt to explain it away via some novel or more fundamental physical mechanism. However, we need not assess the viability of potential future solutions to their Combination Problem here. The simple reason is that if any form of biopsychism can succeed at explaining away the Combination Problem, it should also be theoretically available to the Russellian panpsychist as

⁸³ See also, Reber (2019).

a method to explain away their Combination Problem, Indeed, whatever mechanism is employed to explain the combination of cellular consciousness into multicellular organisms, could in theory also be used by the Russellian panpsychist to explain mental combination, simpliciter (assuming they endorse that meaningful combination only occurs in living things). In this manner, while the Combination Problem *could* be explained away by the biopsychist physicalist, this might also solve the problem for Russellian panpsychism. Furthermore, while Russellian panpsychism would then have solved its main theoretical issue, physicalism (in at least its type-c variant) would still retain its (now even harder) Hard Problem.

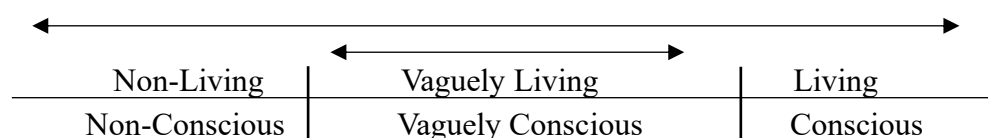
Setting aside the Hard Problem and the Combination Problem, type-b and -c physicalism might also run into a third and independent issue: that ‘life’ is a potentially (linguistically) vague concept, and so cannot signify the required non-arbitrary point for the origin of consciousness (Malaterre, 2024). The philosophical question “What is life?” is relatively absent within practical science. Instead, questions on the origin of life, and the potential existence of life on other planets are much more visible (Morange, 2012: 425). In practice, however, most academics typically revert to working definitions like NASA’s, which asserts that “life is a self-sustained chemical system capable of undergoing Darwinian evolution” (Joyce, 1994: xi-xii; Horowitz & Miller, 1962: 423-459). Such styles of definition typically portray life as a binary construct, such that it is precise and does not admit vague or indeterminate cases: for example, being capable of undergoing Darwinian evolution is a precise necessary condition, and, when paired with the ‘self-sustaining’ condition, appears binary in its application.

While binary definitions regarding life are a useful tool to help classify entities under a series of necessary and sufficient conditions, such definitions typically do not withstand close scientific scrutiny (Malaterre, 2024; Smith et al. 2021). For example, in microbiology, an

increasing number of microscopic entities like reduced microbial symbionts and giant viruses defy any clear classification, indicating a vagueness or borderline instance of life (La Scola et al, 2003; Nakabachi et al, 2006). Similarly, in systems chemistry and synthetic biology, artificial entities have been synthesized that challenge our traditional views of life (Blain & Szostak, 2014: 615-640; Ameta et al, 2021; Otto, 2022: 145-155). Finally, even LUCA has been proposed to have been ‘half-alive’ (Weiss et al, 2016). As a result, more nuanced definitions and approaches to life – grounded in the existence of these vague instances – have emerged which acknowledge that life admits of borderline instances that occupy a sort of grey area between the living and non-living (Malaterre, 2024; Bruylants, et al, 2010: 137-143).

If such reasoning is correct, and life is a vague concept, then life itself cannot signify a non-arbitrary point for the origin of consciousness for the physicalist. If so, then the biopsychist physicalist position would become redundant: life would no longer signify a sufficient point for consciousness to ‘wink in’, as it would itself be formed on a gradual transition from non-living things to living things. For this reason, the physicalist who wishes to find a non-arbitrary point for the origin of consciousness would then have to go would have to posit the origin of consciousness somewhere ‘beneath’ life, and prior to evolution, such that it originates with something non-living and more fundamental than the most basic living things. Clearly, this would again be an unpalatable position for the physicalist, as it would dissolve any significant difference between their thesis and Russellian panpsychism, whilst still retaining their Hard Problem and new Combination Problem. The physicalist might instead still opt to tie consciousness to life, and so concede that consciousness, like life, must also be vague. However, doing so would cause them to fall victim to the aforementioned issues regarding consciousness as vague, such as the potential reduction of physicalism into Russellian panpsychism.

Now, it might be the case that the physicalist can find some non-arbitrary distinction in the physical constitution of those things on the top edge of the determinately not living category from those on the bottom edge of the vaguely not living/living category, and from those on the top edge of the vaguely not living/living category from those on the bottom edge of the determinately living category. Such a non-arbitrary point could then be tied to its equivalent distinction on the aforementioned spectrum of consciousness, so as to provide a non-arbitrary point for the origin of consciousness that still endorses biopsychism, such that:



Unfortunately, this technique will run into the same issues that were presented to the physicalist who proposed consciousness to be vague: any apparently non-arbitrary point between two of our three categories within the spectrum of determinately not life | vaguely not life/life | determinately life will intuitively be of a fundamentally small physical difference; and, if it is the case that some fundamentally small physical difference was the distinguishing factor between two of our categories (and their equivalent categories in the consciousness spectrum) then there must be something about that fundamentally small physical difference (in terms of fundamental parts themselves, or in their arrangement) relevant to the instantiation of life and consciousness. Along these lines, the only way the biopsychist physicalist can make it plausible that such a small difference could make this difference is if he grounds his theory in something similar to panpsychism; the physicalist must again attribute the property of consciousness – or properties relevant to instantiating it when appropriately combined – at the fundamental level.

It might also be the case that life is not vague, and so the physicalist is invulnerable to this final concern. As noted, current definitions of life appear relatively binary, it is just in their application to certain entities that they appear indeterminate. Some commentators think that life is a real natural kind, and indeed, has an exhaustive and exclusive list of properties that we must discover (Bedau & Cleland, 2010). If that is the case, then life is simply not vague. Others consider life to be a property cluster, whereby different definitions propose certain necessary or sufficient properties that an entity must have in order to be classified as living. Within this, one simply needs *enough* of those potential properties within that definition to be classified as living, but typically, how many properties is *enough* is often undecided (Pályi et al. 2002, Popa, 2004). Within this view, once all of the relevant properties are accounted for, and how many are required for something to be ‘alive,’ in principle, life becomes a precise property cluster. Now, if these sorts of reasonings are correct, this final criticism would be removed for the physicalist. I shall leave this matter open. Nonetheless, physicalism will still suffer its inheritance of the Combination Problem, and its worsening of the Hard Problem, and so may still be taken to be less theoretically promising than Russellian panpsychism.

With these aforementioned issues considered, we can see how both variants of physicalism – type-b and type-c – run into some pervasive issues when considered against the evolution of consciousness, such that they must be taken to be less theoretically promising than Russellian panpsychism. We saw how the physicalist has good intuitive reasons to endorse the claim that consciousness is precise. Furthermore, we saw how those who continue to resist claim this will face a particularly strong Sorites Paradox, that either renders consciousness incoherent (such that physicalism cannot provide a satisfactory account to the Mind-Body Problem), or again threatens their theory’s reduction into something similar to Russellian panpsychism. However, we also saw that the physicalist who endorses the precision of consciousnesses will naturally

be led to ‘biopsychism’ and will then run into three dangerous problems: the first, was that the Hard Problem for type-c physicalism gets insurmountably harder; the second, was that both type-b and -c physicalism now possess a similar Combination Problem to Russellian panpsychism; and, the third, was that biopsychist physicalism might itself be potentially unstable – as ‘life’ might be a vague concept – and so the theory may still reduce into something similar to Russellian panpsychism.

At the beginning of this chapter, I stated that we had returned to the theoretical ‘level playing field’ between Russellian panpsychism and physicalism that was first proposed in Chapter 2. Here, I proposed that since both theories have been shown to have respective (and unresolved) epistemic issues, in order to give one or the other theory the advantage, progress must be made into explaining one of their problems away, or a new argument which worsens the problem(s) of the other theory must be provided. This chapter has since shown how questions concerning evolution of consciousness introduce some pervasive issues for physicalism: however the physicalist proceeds with regards to answering the question of *when* it is that consciousness ‘pops up’, their theory threatens to either reduce into Russellian monism, inherit its Combination Problem, worsen its Hard Problem, or indeed, result in some combination of all of the above. In this manner, this chapter should serve to prove that the theoretical level playing field is no more: if physicalism must reduce into Russellian monism to provide an answer to the argument from evolution, then it ceases to be a distinct theory to that which it is competing against; and, if it maintains all of its own issues whilst inheriting the most damaging argument against Russellian monism, or if its own Hard Problem becomes insurmountably ‘harder,’ then physicalism must be taken to be less theoretically promising than Russellian monism, all other things considered.

For the sake of fairness with regards to these claims, we should also assess how Russellian panpsychism fares against the same Argument from Evolution. It is no good claiming that physicalism falls victim to the Argument from Evolution, and so is less theoretically promising than Russellian panpsychism if the latter has not been compared against the same charge. However, as I'm sure the reader will already deduce, the worry of *when* consciousness popped up within our evolutionary trajectory does not truly present any form of argument against Russellian panpsychism, and so the argument itself fails to materialise.

Since quiddities are the non-structural, non-relational, categorical properties which underlie and characterise the micro-level entities posited by physics, and since for the Russellian panpsychist quiddities are phenomenal properties, then phenomenal properties underlie and characterise the micro-level entities posited by physics in its characterisation of matter. In this manner, the Russellian panpsychist doesn't have to entertain questions about vagueness or precision with regards to consciousness: consciousness is the categorical property of fundamental entities, and so it resides within every composite thing in the precise sense, regardless of whether or not it is combined into a singular macro-subject. Similarly, the Russellian panpsychist doesn't have to give an account as to *when* consciousness pops up within our evolutionary trajectory, since consciousness comes before evolution. With this in mind, and with regards to the Argument from Evolution itself, it should be clear that Russellian panpsychism has the upper hand: it inherits no further issues from the argument over that which it already suffered, and crucially, that same issue is now also suffered by physicalism. In this manner, the Argument from Evolution shifts the theoretical playing field in favour of Russellian panpsychism.

While the worry of *when* consciousness popped up may not present any issue to the Russellian panpsychist, they must still resolve the issue of accounting for *how* micro-experience can combine into macro-experience. This is just to say that while physicalism has now made its Hard Problem ‘harder,’ and has inherited Russellian panpsychism’s Combination Problem, the Russellian panpsychist still must resolve the Combination Problem if it is to be internally coherent. With this in mind, the question of when and how do these quiddities combine to make complex consciousness remains an issue that is shared with the physicalist and is yet to be resolved. In part, this will be the focus of the next section.

5.4 ~ Further Considerations

As stated within the introduction to this chapter, aside from providing an argument against physicalism, and for Russellian panpsychism, my Argument from Evolution is intended to serve two functions. First, it was proposed as a safeguard for Russellian panpsychism, such that even with the Combination Problem looming, Russellian panpsychism can still provide a more attractive theory of consciousness than physicalism. Second, I suggested that potential and future responses to the Combination Problem could build from the logic of the Argument from Evolution, such that this chapter can serve as a foundation for later authors to respond to the problem effectively. While the former function has been satisfied, the latter still needs to be defended. With this in mind, let us turn to the question of how my Argument from Evolution can inform future responses to the Combination Problem. Crucially, these potential forthcoming responses and contextualisation are solely intended as interesting avenues for future research, rather than constituting specific arguments in-and-of-themselves.

5.4.1 Prospective Responses to the Combination Problem

Russellian panpsychism has been argued to be immune to the Argument from Evolution. Since quiddities (as phenomenal properties) are the non-structural, non-relational, categorical properties which underlie and characterise the micro-level entities posited by physics, consciousness resides within every simple and composite thing in the precise sense; simple consciousness originates prior to evolution, and so the evolution of living things already has consciousness entailed within that process. That being said, and as noted, the Combination Problem still exists, and so the Russellian panpsychist must still answer the question of when and how do quiddities combine in a relevant sense to make macro-consciousness. While this consideration is independent of Russellian panpsychism's status as the best available account of consciousness, it remains a necessary hurdle that Russellian panpsychism must resolve if they are to establish their theory as the *correct* account. With this in mind, let us speculate on potential avenues for future areas of research and argumentation in this regard.

Now, within any response to the Combination Problem, the Russellian panpsychist must answer the question of *when* micro-consciousness began to combine into more complex forms of macro-consciousness; the reality of a combination principle in nature appears to necessitate that one can draw a distinction between complex conscious things, and simple conscious things. In this manner, and in a similar sense as to that which was requested to physicalism, the Russellian panpsychist should be able to point to a time in which the complex consciousness of certain entities combined from the simple consciousness of fundamental parts. To begin structuring potential avenues of response to this question, we could refer back to the first two stages of James and Clifford's argument, which stated that:

- A. Evolution can only assemble pre-existing properties into further complex forms.

- B. Given A, evolution can only assemble pre-existing consciousness properties into further complex forms.

Now, since it is the case that evolution can only assemble pre-existing properties into further complex forms, the Russellian panpsychist has a number of potential options as to where the relevant combination could broadly occur:

Option 1: the combination of micro-consciousness occurred prior to evolution, such that complex consciousness already existed at the origin of living things, and evolution then continued to assemble this property into further forms.

Option 2: the evolution of complex consciousness occurred at some point within the broader evolution of living things.

Option 3: the combination of micro-consciousness coincided with the origin of life, such that complex combination exists in all living things, to varying degrees of complexity.

To endorse Option 1 would be to claim that certain or perhaps all inanimate, non-living composite objects possess complex consciousness; if complex consciousness exists prior to evolution, it originates at a level of reality that is ‘lower’ than living things. In this manner, such a principle would not originate in, nor be dependent upon evolution. For example, it might be a principle of a similar fashion to that which entails that the mass of any composite object is the combination of the mass of its parts. Now, from the offset, this option is probably the safest for the Russellian panpsychist. As noted, it might be that ‘life’ is a linguistically vague concept. If so, then life may fail to signify a non-arbitrary point for which mental combination could occur. At the very least, one may introduce complications regarding mental combination being vague that opens oneself up to concerns similar to those proposed to the physicalist.

Through adopting Option 1, and by making mental combination a fundamental mechanism, we can bypass these concerns entirely.⁸⁴

Luke Roelofs is perhaps the most prevalent supporter of this option. In their (2019) book *Combining Minds*, Roelofs proposes that subjects are substrates, not personas, that composites are aggregates, nor structure-specific wholes, that consciousness is fundamental, and that the combination of consciousness is primitive (Roelofs, 2019: 70). In this manner, Roelofs proposes that consciousness is present in any aggregate in the same fashion as the consciousness is present in its parts.

To account for *how* this combination occurs, Roelofs proposes two principles. The first is through ‘Experience Inheritance,’ whereby whenever a part of aggregate *x* undergoes an experience (instantiates a phenomenal property), *x* undergoes that same experience. In this manner, if one particle is undergoing experience A, and another is undergoing experience B, the aggregate of the two is undergoing both A and B. Crucially, this does not involve any duplication of experiences, but rather a sharing of experience between part and whole (ibid: 79-80). The second principle is the ‘Micro-Unity Hypothesis,’ whereby the inner nature of one, some or all of the fundamental physical relations is phenomenal unity; when two micro-subjects are related in the relevant way, their experiences become unified, establishing a composite experience that subsumes them (ibid: 80). In this manner, the Micro-Unity Hypothesis is akin to any basic physical relation; if any two quiddities interact or are physically related (no matter how small that relation or interaction), then they enjoy phenomenal unity with each other (ibid: 94).

⁸⁴ It should be noted that if my ‘backup argument’ regarding Sorites paradoxes holds, and life is indeed vague, then life might not itself be a real property. If this is the case, then even more work must be done to satisfy Option 2 and 3. We will consider avenues in this regard shortly.

While Roelofs provides a compelling and rigorous account for the viability of Option 1, and it probably presents the safest avenue for Russellian panpsychism with regards to resolving its Combination Problem, Roelofs admits that his account produces some contentious and extreme consequences. For example, by virtue of their specific combination principles, one is led to the contentious conclusion that one's socks might possess complex consciousness in a meaningful sense (Roelofs, 2021: 120). Furthermore, one is led to the (perhaps even more contentious) conclusion that the entire universe is phenomenally unified, such that my consciousness, the readers consciousness, and the consciousness of our socks all participate in a larger, unified, 'universal mind' (Roelofs, 2019: 95-97). As Barry Dainton puts it, given "every particle composing the planet Earth is linked (directly or indirectly) to every other by a chain of physical connections ... the entire planet will consist of a single fully unified consciousness" (Dainton, 2011: 255). This is indeed a contentious conclusion to have to draw, and while Roelofs is happy to concede such, others may not. With this in mind, let us look to Option 2 and 3 as future alternatives for the Russellian panpsychist.

Now, both Options 2 and 3 necessitate that the combination of micro-consciousness occurred at the origin of, or as subsequent to life. In this manner, both necessitate that the combination principle is intimately related to those principles that govern living organisms or evolution itself. For example, with regards to *when* it is that simple consciousness began combining into complex consciousness, Option 2 necessitates that we can signify some point within the general evolutionary development of living things where combination began to occur. Option 3, on the other hand, ties combination to the origin of life, such that the complex combination began with the first living things and has since evolved into varying degrees of further complexity.

Crucially, both options necessitate that *only* living things possess complex consciousness, such that no non-living things enjoy any combination of the simple consciousness of their parts.

One might be immediately concerned that through intimately relating mental combination to life, one will have to point to some small difference that separates things with mentally combined consciousness from those that don't. In a similar manner to that which was proposed against physicalism, one might allege that it is just entirely implausible to think that any tiny difference could make such a difference. Plausibility aside, in the case against physicalism, I proposed that any fundamentally small difference which could signify the origin of consciousness would threaten physicalism's reduction into panpsychism. Conversely, there is no theory X such that this small difference relevant to mental combination would threaten a reduction of panpsychism into theory X. In this regard, Russellian panpsychism clearly has advantage over physicalism. Furthermore, and in the case of Russellian panpsychism, this is an epistemic gap between two conceptually similar things – micro-consciousness, and (combined) macro-consciousness – rather than between two conceptually dissimilar things – physical things, and conscious things. In this manner, it is certainly more plausible in the case of Russellian panpsychism that a small difference could account for the change, than it is for physicalism.

Now, like Option 1, both Option 2 and 3 signify potential points as to *when* complex consciousness began to occur. However, unlike Option 1, both necessitate that their potential combination principle is intimately related to life, such that they draw similarities to the biopsychism. On the one hand, these options might present a slightly more intuitively compelling picture than Option 1. Indeed, at least in our experience, life and macro-experience appear to be intimately linked: all the things that we want to say are relevantly conscious are

living, and we are yet to come across any non-living thing that we would intuitively wish to claim has any macro-experience. Furthermore, our consciousness clearly evolved as we evolved, such that there are clear links between the evolution of consciousness, and the evolution of living things. In this manner, it is potentially intuitive that any prospective combination principle would likely be intimately related or tied to life and evolution. On the other hand, however, by intimately relating any potential combination mechanism to ‘life,’ Option 2 and 3 might open themselves up to the potential worry that ‘life’ might be a vague concept or property.

There are two options for the Russellian panpsychist in this regard. First, they could endorse that life is precise, such that life is a natural kind, or a precise property cluster, which entails an exhaustive and exclusive list of properties that entities either satisfy or they do not (Bedau & Cleland, 2010; Popa, 2004). Now, this would remove the third horn of argumentation against our biopsychist physicalist position. However, and as noted, this need not present an issue to our case here; biopsychist physicalism will still have inherited the Combination Problem and have worsened its Hard Problem, and so Russellian panpsychism can still be theoretically preferred. However, the Russellian panpsychist might instead choose to endorse that life is vague and so retain this issue for physicalism. If they choose this option, they must account for whether or how mental combination is also potentially vague.

Let us now speculate on potential combination principles for each option, with consideration to the fact that life may or may not be precise. As before, these principles are solely intended as prospective avenues for future research, as opposed to being attempted resolutions in-and-of-themselves. Now, if life is precise, Option 2 has a clear avenue for a potential combination principle. Recall, this option necessitated that any combination principle is intimately linked to

the principles that govern evolution itself. Now, as noted in P.1, evolution assembles pre-existing properties into further complex forms. In this manner, evolution *is* already a genuine and existing combination principle; it accounts for why simple living forms have developed into more complex living forms. Accordingly, the commentator who endorses the precision of life, and chooses to pursue Option 2 already has an existing combination principle to build from, such that they need only modify the existing principles of evolution.

For example, our Russellian panpsychist could propose that living entities have evolved to enjoy a state of unity between micro-experiential parts that mirrors the evolution of, for example, reproduction, homeostasis, and self-organisation. Indeed, the persistence of a living thing has been likened by Jonathon Miller as an achievement of the same order as that of a fountain, such that: “the material from which such an object is made is constitutionally unstable; it can maintain its configuration only by flowing through a system which is capable of reorganising and renewing the configuration from one moment to the next. But the engine which keeps a fountain aloft exists independently of the watery form for which it is responsible, whereas the engine which supports and maintains the form of a living organism is an inherent part of its characteristic structure” (Miller, 1978: 140-141).

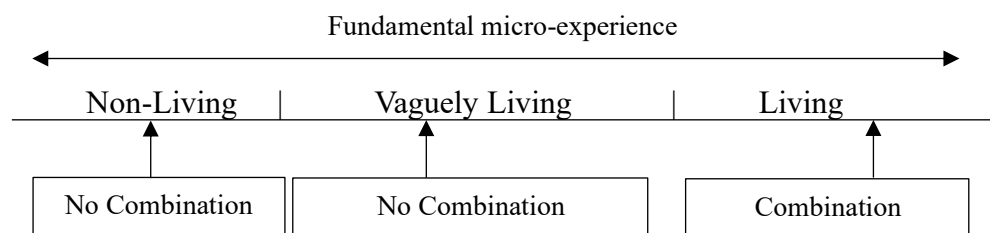
The Russellian panpsychist might propose that the physical relations of dependence, organisation, and collaboration that exist between our parts – and constitute our ‘engine’ – simply function better when the micro-experiences of those parts are shared or ‘communicated’ with each other – and by consequence, the whole. One could posit that a principle similar to Roelofs’ ‘Experience Inheritance’ exists conditionally within living systems, and so once the micro-experiences of those parts began to ‘stack’ in the relevant matter, the process continued to be passed on as an evolutionarily beneficial trait and evolve accordingly into the form in

which we enjoy today. Crucially, this is merely speculation and warrants a far more robust analyses than that which can be afforded here. Nonetheless, we can see that there are clear future avenues for research by the Russellian panpsychist in this regard. Indeed, in tying future combination principles to evolution, and by proposing that macro-consciousness only exists in living things, one would surely also resolve the Boundary Problem: things outside of living unified systems are not parts of that unified system, and so the unification of micro-experiential parts does not extend to them.

I propose there is a clear sense in which Option 2 also presents an avenue for the Russellian panpsychist even if life is vague. Where the physicalist needed life to signify a non-arbitrary point for the origin of consciousness, the panpsychist already has precise consciousness as a fundamental and ubiquitous reality of reality, and so life being vague does not add vagueness to consciousness. Instead, if life is vague, then the proponent of Option 2 has two potentially viable options. The first, is to propose that although life is vague, mental combination is precise and occurs *only* in those things that are determinately living within our spectrum of classification. In this manner, one can again posit that a principle similar to Roelofs' 'Experience Inheritance' evolved as a shared property within (only) living systems, such that once the micro-experiences of those parts began to 'cluster' or 'stack' in the relevant matter, the process continued to be passed on as an evolutionarily beneficial trait and evolve accordingly into the form in which we enjoy today.

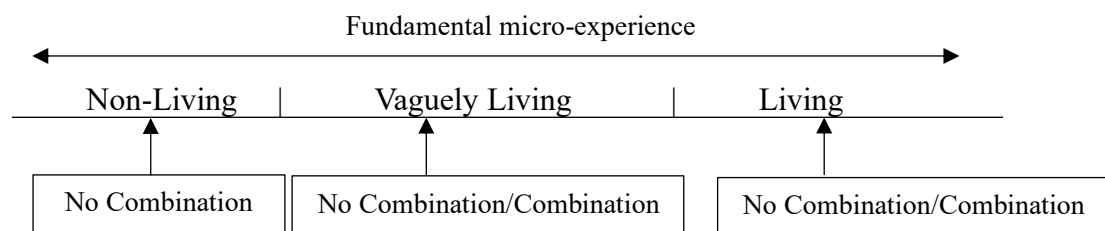
As noted, Option 2 only necessitates that the combination of micro-consciousness occurred *somewhere* within the process of evolution, such that the evolution of complex consciousness occurred at some point within the broader evolution of living things, as a beneficial evolutionary trait. By virtue of this, it might be that only determinately living things enjoy any

form of evolved mental combination, and no vaguely living thing does (since they have not evolved, or at least, followed an evolutionary lineage relevant to the evolution of mental combination). If so, the proponent of Option 2 can avoid speculation entirely about vague mental combination, and utilise the above potential mechanism associated with life being precise. We could illustrate this as follows:



The second option is to endorse that although life may be vague, it is only vague insofar as the property cluster has unbounded edges. In this manner, it is somewhat arbitrary as to where we should draw our boundary with regards to what properties constitute life. Crucially, while the property cluster may be linguistically vague in this way, there are many properties within the cluster which are typically understood to be precise, for example, the property of reproduction, or growth. Now, it can be vague as to whether particular entities in possession of such properties are living. For example, a virus is typically classed as non-living, or sometimes as a vague or borderline case, by virtue of its having – to name few – the necessary and precise property of reproduction, but not the necessary and precise property of growth (Bedau, 1996; 1998). In this manner, we can see how ‘life’ can be linguistically vague – such that it represents a useful fiction with regards to describing which entities have such and such properties – without the particular properties it encompasses also being necessarily vague themselves.

Utilising this reasoning, the Russellian panpsychist might simply choose to propose that mental combination is a determinate property within the indeterminate property cluster of ‘life;’ things either have that precise property, or they do not. Those that do hold that property, and which can evolve, can then enjoy the evolution of that property; once the micro-experiences of those parts began to ‘cluster’ or ‘stack’ in the relevant matter, the process continued to be passed on as an evolutionarily beneficial trait and evolve accordingly into the form in which we enjoy today. In making this move, one can uphold that consciousness and mental combination are both precise properties, and so avoid the aforementioned issues suffered by physicalism regarding vagueness. We can illustrate this as follows:



With these options set up, let us turn to Option 3. To begin, let us again assume that life is precise. To support the Russellian panpsychist’s future inquiry into how meaningful combination could occur at the point of life, we can appeal to van Inwagen’s account of parts and wholes. Per van Inwagen (and where the parenthetical disjunct is included to secure the reflexivity of parthood):

$(\exists y \text{ the } xs \text{ compose } y)$ if and only if

the activity of the xs constitutes a life (or there is only one of the xs).

(van Inwagen, 1990: 82).

Here, van Inwagen suggests that only living things and simples exist, such that only self-

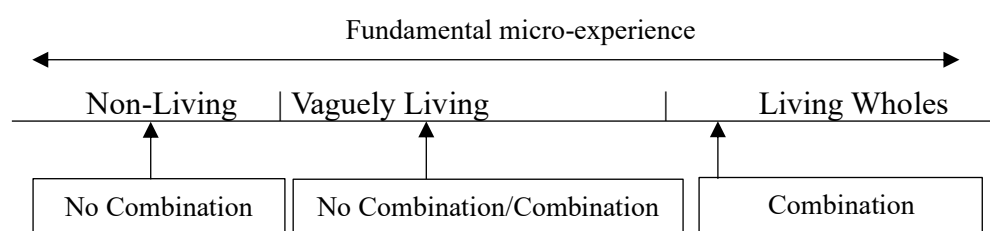
maintaining things are composites (wholes), only living things are self-maintaining, and so only living things are composites (wholes). In this manner, things like tables and chairs are not wholes, but rather simply aggregates of simples arranged in a certain way; there is no additional ‘table-ness’ that makes a table a unified entity. Conversely, living things are true composites or wholes, because they have a biological unity allows it to persist, grow, maintain itself, and function as one (ibid: 81-97). Following van Inwagen, the Russellian panpsychist could maintain that only living things are wholes, such that only living things are true composites of the simples (or fundamental parts) which compose them. They could then tie this to a combination principle such as Roelofs ‘Experience Inheritance’, whereby the living whole enjoys composite experience of its parts (Roelofs, 2019: 79-80). Crucially, since only living things are wholes, only living things enjoy whole-parthood relations of this form. Furthermore, since only individual living things will then enjoy Experience Inheritance, one should avoid the Boundary Problem of our Experience Inheritance extending beyond individual living things.

Another potential avenue would be for the Russellian panpsychist to uphold van Inwagen’s claim, and then modify the grounding relation within this thesis, such that wholes ground their parts *by subsumption*, as opposed to those parts grounding by analysis. As introduced in Section 2.2.3, we can say that some entity X *grounds by subsumption* entity Y iff: (1) X grounds Y, and (2) X is a unity of which Y is an aspect (Goff, 2017: 221).⁸⁵ Through grounding by subsumption, the Russellian panpsychist could propose that all living things (as wholes) ground their parts by subsumption, such that all and only living things enjoy the unification of the micro-experience of their parts. Since only living things would be wholes, the subsumption

⁸⁵ While one might be worried for our grounding relation that the parts pre-exist the whole, we can say here that prior to their subsumption, they are only simples, and so their pre-existence need not ‘muddle the water’ with regards to grounding. Truths about those simples *when* subsumed by the whole *then* exist in virtue of truths about that whole.

relation would only occur between those things that are living, and their parts, and thus again prevent any boundary issues regarding combination principles: mental combination would not extend any further than individual living wholes.

I propose there is also a clear sense in which Option 3 presents an avenue for the Russellian panpsychist even if life is vague. Indeed, van Inwagen himself concedes that there is vagueness as to when life begins, or more precisely, there can be vagueness about the parthood of wholes (van Inwagen, 1990: 20). As with Option 2, I don't believe this needs to signify a worry for the Russellian panpsychist. For example, and building from the second option for Option 2, we can say that although life may be vague, it is again only vague insofar as the property cluster has unbounded edges. As above, perhaps mental combination is a determinate property within this indeterminate cluster, but unlike before, it is a *necessary* property to be determinately living, in the same way that reproduction, growth, evolution, and so on, are typically deemed to be (Bedau, 1996; 1998, Joyce, 1994). In this manner, we can say that while the property cluster 'life' may be indeterminate, all living things are wholes, and – utilising van Inwagen's part/whole relations, or Experience Inheritance – each possesses the necessary property of mental combination of its parts. Crucially, doing so would maintain both the precision of consciousness and of mental combination, and so avoid any of the aforementioned issues suffered by physicalism. With regards to the property of combination, we can illustrate this as follows:



As noted, these options are merely representative of potential future research that the Russellian panpsychist may undergo with regards to resolving their Combination Problem. Options 1 – 3 each appeal to the logic of the Argument from Evolution so as to establish potential answers to the question of ‘when’ consciousness began to combine in a meaningful sense. I have then explored different avenues for each with regards to the question of ‘how?’. While each option and its answers require much more extensive work, each still represents a possible solution in principle to the Combination Problem. With this in mind, we have further bolstered the claim that Russellian panpsychism should be taken to be theoretically preferable to physicalism, such that it provides the best available account with regards to resolving the Mind-Body Problem.

5.4.2 Further Contextualisation

I have argued that physicalism runs into some pervasive issues when questioning how it is that consciousness fits into the evolutionary model. I have argued similarly that Russellian panpsychism is immune to such worries, and that it provides the best available account with regards to solving the Mind-Body Problem. In this sub-section, I will briefly provide some potential context for these issues. In particular, I shall explore potential evidence that some of the most dominant proponents and developers of the evolutionary model – namely, Charles Darwin, George Romanes, and Sewall Wright – believed in the biopsychist thesis, and thought it was required if consciousness is to be reconciled with evolution. I will then explore whether there is evidence for them grounding this thesis in anything further, for example, in physicalism, or panpsychism.

I shall unpack the shared belief of Darwin, Romanes, and Wright, that mental evolution requires consciousness to exist in all living things by virtue of the principles of organic evolution and

our shared unity of decent. While Darwin does not explicitly ground this thesis in any metaphysical account of consciousness, I shall propose that Romanes and Wright ground theirs in panpsychism. In this manner, we can say that certain predominant evolutionists believed that mental evolution requires biopsychism, and that this biopsychism requires panpsychism in order to be metaphysically complete. Furthermore, and as we shall see in the case of Wright, this panpsychism is built into and appears to have logically motivated crucial aspects of his science. As before, this analysis is intended to be representative of future research with regards to the Argument from Evolution.

These considerations may help to shed light on why the physicalist who concerns himself with the evolution of consciousness will find himself adopting biopsychism, and also why his account risks reducing into panpsychism. Similarly, such considerations may also shed light on why panpsychism appears immune to the evolutionary worries suffered by physicalism. This is not to say that such context proves physicalism to be false, or indeed, panpsychism to be true. Instead, this analysis serves only to help us recognise the assumptions that might be contained within the evolutionary model; clearly, any assumptions inputted into a theory must be explicated if we are to truly understand how its developers intended for that theory to be used, and, when explicated so in this case, may prove useful to concerns about consciousness and evolution. In this manner, this section might help explicate the context relevant for the issues suffered by physicalism in the previous section, and indeed, might help explicate the context for panpsychism appearing immune to such considerations.

Now, the question of “how does consciousness commence?” is one which Darwin did indeed consider, and occasionally wrote on (Darwin, 1838: 35). That being said, most of the available comments have only been published postmortem from a series of notebooks he kept, and none

of which directly support any theory of consciousness. The notebooks only reveal some ‘ferments’ of ideas from which his more mature works grew, and such interests in consciousness and comparative psychology are only small aspects of his overall purpose (Smith, 1978: 257). This is not to say that they were not important questions to Darwin, but rather only that it is unsurprising they failed to receive the full consideration that he afforded to that material more central to his theory. With that in mind, my analysis below does not purport to claim Darwin to hold any particular view, but rather serves only to collect his explicit comments and questions on consciousness so that we might better understand those inclinations. In this light, and after considering Darwin, we shall turn to his young contemporary and friend, George Romanes – the recipient and developer of the majority of Darwin’s unpublished work – whose explicit comments on such matters might better elucidate Darwin’s less explicit position.

Darwin clearly understood humans and animals to be ‘netted together’ through their shared line of descent. In his B notebook, for example, he writes: “Animals our fellow brethren in pain, disease death and suffering and famine ... our companion in our amusements, they may partake, from our origin in one common ancestor we may all be netted together.” (Darwin, 1938: B.232). Darwin believes that it is largely human arrogance which prevents us from treating animals as our equals. In notebook C, for example, he asserts: “Man in his arrogance thinks himself a great work worth the interposition of a Deity. More humble and I believe truer to consider him created from animals.” (ibid: C.196-197). With regards to those particular properties we share – alongside the aforementioned abilities to feel pain and suffer – he states that “to see a puppy playing cannot doubt that they have free-will,” and hence consciousness (ibid: M.72; OUN.25a). He clarifies further, that if “all animals, then an oyster has and a polype” also (ibid: OUN.25a). In the same manner as we are unified in ‘type’ by our unity of

descent then, Darwin extends that like ourselves, all animals have consciousness (Darwin, 1859: 233).⁸⁶ Accordingly, we can say that for Darwin, humans and all animals are conscious.

Darwin clearly views all animals as conscious beings. The prudent question is then how far throughout the living world does he think consciousness extends? While he provides no clear or obvious answer to this question, there is textual evidence to support the claim that he believes it to extend to at least all living things, and thus to hold a view akin to contemporary biopsychism (Smith, 1978: 256).⁸⁷ For example, he asks rhetorically “[c]an insects live with no more consciousness than our intestines have?” (Darwin, 1838: OUN.35). Within the broader context of this quotation, we can infer that he believes lower-level animals such as insects to be endowed with consciousness. Furthermore, and regarding plants, he states that “it is hardly an exaggeration to say that the root tip of the radical [primary root] . . . acts [like] the brain of one of the lower animals . . . receiving impressions from the sense organs and directing the several movements” (Darwin, 1880: 573). Similarly, regarding William Kirby’s 1835 *Bridgewater Treatise*, where Kirby conceived that the development of plant seed-dispersal mechanisms was an instance of instinct operating, Darwin writes that their instinct is evidence of “Reason, Will, [and] Consciousness” (Darwin, 1838: OUN.36). In this manner, it seems that not only does Darwin believe animals to be endowed with consciousness, but that he also might believe it to extend to many other living organisms.

Now, if we are to understand Darwinian reductionism and his theory of unity by descent, we can see that he wishes to eliminate any deep drawn demarcation in type between humans and

⁸⁶ Indeed, such reasoning appears identical to that which motivated Clifford and James’s initial argument from evolution.

⁸⁷ I am aware of the possible anachronism of attributing to Darwin any contemporary viewpoint. My comparison of his comments against these concepts is intended only to better understand his view, rather than propose they should be directly attributed to him.

any other living thing. Furthermore, and as we have seen, his explicit comments suggest that he believes that consciousness exists in animals and many other living organisms. Following his wish to eliminate any demarcation in type within living things, alongside his specific claims about mammal, animal, and plant consciousness, there seems to be some reasonable evidence to suggest that Darwin believed consciousness to exist in at least all living things, by virtue of their unity by descent. The obvious next question is then, does he believe that consciousness extends *beyond* living things?

Darwin clearly appreciates that consciousness can be divided within living things. For example, he states that “A Planaria [a type of free-living flatworm] must be looked at as an animal, with consciousness, it choosing food – crawling from light – yet we can split Planaria into three animals and this consciousness becomes multiplied” (ibid: OUN.16). In this manner, Darwin appears to suggest – *a la* panpsychism – that living conscious organisms can be divided into smaller conscious things, where each of which retains a consciousness that is no less than the original organism. However, he found more fundamental questions about the origin of consciousness – such as how far this division can occur – to be extremely perplexing. Indeed, following up from comments on the similarities between chemical attraction amongst elements and thoughts, he goes on to ask, but “What is matter? The whole thing a mystery.” (ibid: OUN.37). In this light, his notebooks provide no clear or explicit answer to these more fundamental questions, aside from some tangential similarity to the questions proposed by Russellian panpsychism regarding the universe’s fundamental constituents (Darwin, 1838: 35).

Unfortunately, we find later in folio 2 to his *Natural Selection* that he has given up on such questions entirely, stating that “we are no more concerned with the first origin of the senses and the various faculties of the mind, than we are with the first origin of life” (Darwin, 1975:

467). This same sentiment is shared in his *Descent of Man*, where he writes: “in what manner the mental powers were first developed in the lowest organisms is as hopeless an enquiry as how life itself first originated. These are problems for the distant future, if ever they are to be solved by man” (Darwin, 1901: 100). In this manner, although we can say that Darwin certainly considered many questions regarding the extension and origin of consciousness in the natural world and held a view with a form potentially similar to contemporary biopsychism, we can extract no definitive metaphysical perspective on consciousness from his explicit remarks.

In order to help better understand Darwin’s potential position on such matters, we should turn to his younger contemporary and friend George Romanes. It is well known that Darwin handed over his notes, clippings, and manuscript on comparative psychology to him, and that Romanes heavily wove this material into his book *Mental Evolution in Animals* (Smith, 1978: 258). In this manner, Romanes’ view on these questions might be instructive of Darwin’s own unpublished and less explicit position. Crucially, this is not to say that Darwin actually held the same views as Romanes, but rather simply that Romanes was obviously influenced by Darwin. At best, Romanes’ explicit position might help inform us of Darwin’s implicit position, and at worst, it will simply inform us of the philosophical position held by another dominant and crucial developer of our current evolutionary model.

In his *Mental Evolution in Animals*, Romanes writes that “if the doctrine of Organic Evolution is accepted, it carries with it, as a necessary corollary, the doctrine of Mental Evolution” (Romanes, 1883: 8). Clearly then, Romanes was concerned with *how it is* that consciousness could have evolved, and how we should fit this explanation into organic evolution. He continues, “starting with what I know subjectively about the operations of my own individual mind, and of the activities which in my own organism these operations seem to prompt, I

proceed by analogy to infer from the observable activities displayed by other organisms, the fact that certain mental operations underlie or accompany these activities” (ibid: 16). Shortly after, he describes that all organisms display activity that is indicative of choice, and that the decisions of which are unpredictable to an outside observer. Crucially, he concludes that this unpredictability of behaviour is indicative of mentality (ibid: 17). These comments match Darwin’s understanding that all other living things are conscious and draws comparisons in its formulation to many of Darwin’s explicit comments on such matters, and indeed, in his letters on the topic to Romanes himself (Darwin, 1903: 213). In this light, we say that Romanes – like Darwin – probably held a distinctly biopsychist position.

Now, unlike Darwin, after his consideration of the necessary and sufficient signs of consciousness in organisms, Romanes does proceed to comment on the metaphysical basis of consciousness itself. In particular, he refers to the writings of William Clifford, one of our initial formulators of the Argument from Evolution, and indeed, a prominent panpsychist. For example, in his essay, “The Fallacy of Materialism,” Romanes quotes and explicitly agrees with the panpsychist principles of Clifford (Romanes, 1882: 871-888). In particular, he draws upon the following passage from Clifford’s essay ‘On the Nature of Things-in-Themselves’ (1878): “Mind-stuff is the reality which we perceive as matter ... A moving molecule of inorganic matter does not possess mind or consciousness, but it possesses a small piece of mind-stuff. When molecules are so combined together as to form the film on the under side of a jelly fish, the elements of mind-stuff which go along with them are so combined as to form the faint beginnings of sentience... When matter takes the complex form of a living human brain, the corresponding mind-stuff takes the form of a human consciousness, having intelligence and volition” (Romanes, 1882: 871-888).

We can say then that unlike Darwin, Romanes explicitly grounds his metaphysics of consciousness in a form that is akin to the panpsychism expounded in this thesis. Mental combination is then proposed to occur at the level of ‘life,’ such that the only things with complex consciousness are living things. In this manner, Romanes resolves the explicit confusion Darwin held about the origins of consciousness, in such a way that the evolution of consciousness can now be tracked throughout living things, and indeed, to its fundamental grounds. Whether Romanes’ explicit metaphysics about consciousness was shared by Darwin, is beyond available analysis. However, it clearly began its journey through the work of Darwin and sought to resolve the aforementioned metaphysical issues that he faced on such matters.

Moving to another crucial but more contemporary developer of our theory of evolution – Sewall Wright – we see the same inclinations as Darwin and Romanes again. Wright (1889-1988) founded his scientific career on the special nature of complex gene systems. He designed the popular statistical technique of ‘path analysis,’ and developed a model of organic population flux, that, combined, has become later known under the rubric of ‘theoretical population genetics.’ Crucially, while Wright provided some of most important developments to our evolutionary theory of the last century, according to M.J.S Hodge (1992) and David Steffes (2006), he also held an explicitly panpsychist philosophy. I shall argue further for this sentiment below, before suggesting that this philosophy may have directly influenced and grounded his science. In this light, I shall conclude that it would be unwise of us to discount the particular philosophical views held by dominant formulators of the evolutionary model from our understanding of that model.

Now, like Romanes, Wright’s personal philosophy was also heavily influenced by Clifford (Steffes, 2006). In particular, he held the view that his contemporary physicalist and vitalist

philosophers were wrong; the material conditions upon which physics, chemistry, and biology rely are ultimately just the language in which complex conscious systems interact with each other (Wright, 1975: 80; c.1985: 7-8). For example, he proposed that “[w]e must postulate common consciousness as arising from interactions among elementary physical entities within an atom, or among atoms and electrons within a molecule, or among molecules within a living cell as well as among cells within a multicellular organism in carrying through the concept of [psycho-physical] identism. We should again recognise that the external aspect of a stream of consciousness is not matter defined by mass and occupancy of space so much as by the association of action [energy]. We must suppose that wherever there is physical action, this is the external aspect of at least a flicker of consciousness.” (Wright, c.1985: 10). In this manner, it appears uncontentious that Wright held a distinctly panpsychist view.

In addition to identifying any physical action with ‘flickers’ of consciousness, Wright also believed that any tightly knit system of complex activity represented ‘other minds,’ located on hierarchical levels representative of their degree of conscious sophistication (ibid: 351). For example, Wright recognised that genes are the physical determinant of the animal phenotype, and which operate with great regularity. However, he also recognised that the special nature of genetic systems adds a degree of novelty and unpredictability to the given phenotype, which inspired him to view genes as conscious organisms in-and-of-themselves (Provine, 1986: 83-88; Steffes, 2006: 335, 351). In this light, Wright appears to believe that all living things, and indeed, many non-living things are conscious; living things occupy the higher hierarchical levels representative of more complex consciousness, and those on the lower levels possess more simple consciousness representative of their lower level of complexity of action.

Crucially, much of the historical scholarship on Wright solely focuses on his population genetics, and little has been said on his reasons – philosophical or not – for integrating statistical with experimental biology (Steffes, 2006: 328). According to William Provine – Wright’s biographer and seminal scholar on the origin of population genetics – “Wright’s early views on philosophy... determined only his later views on philosophy, not his work in physiological genetics or evolutionary theory” (Provine, 1986: 95). More recently, however, many contemporary authors have since concluded that Wright’s philosophy heavily influenced his work in genetics and evolutionary biology (Hodge, 1992: 231-293; Steffes, 2006, Ruse, 2004: 131-150). For example, both Hodge and Steffes’ studies concluded that Wright’s panpsychism heavily influenced his work, while Ruse’s account focusses on his broader philosophical commitments. Nonetheless, each author proposes that Provine was incorrect to conclude that Wright’s philosophy had no discernible effect upon his science.

I agree that Wright’s philosophy appears to have had a heavy influence on his work on evolution, and in a sense which is far more active than that which was afforded to Romanes. Wright believed that consciousness exists on many different hierarchical levels of nature, and indeed, throughout all of physical reality, and there is a clear sense in which this appears to have actually motivated his development of path analysis. Firstly, from Wright’s perspective, the conscious entities which determine nature’s organisational processes – such as genes, cells, and so on – give rise to both regular and irregular phenomena in the eyes of the observing scientist. Crucially, for Wright, these irregular phenomena or novelties were indicative of interaction between levels of consciousness in nature (Steffes, 2006: 357). In this manner, the more complex and so conscious an organic system is, the greater the novelty that results (Wright, 1953: 16-17). Secondly, for the scientist to assess the complex interactions responsible for these ordered and novel phenomena, one needs techniques for analysing causation and

correlation (viz, for analysing regular patterns, and irregular patterns, respectively). Crucially, for Wright, the cause of the novel phenomena is consciousness!

So as to create a technique to assess the complex interactions responsible for these ordered and novel phenomena, Wright developed his path analysis. In this sense, the unprecedented importance he places upon finding both causal and correlative relationships was logically proceeded by his panpsychism. Furthermore, since Wright believed that the cause of the novelty in question is consciousness, his path analysis acts – in part – to assess what he believes to be the interaction of levels of consciousness across nature. Not only do we then appear to have another distinctly panpsychist developer of evolution, but we appear to also find one who thought it necessary to weave such philosophy into his science.

With this in mind, we should be unwilling to discount the particular philosophical views held by Darwin and Romanes from their presentation of evolution. In a similar way to Wright, it might be that their views held a crucial role in the development of the theory itself. Indeed, if each of these developers of the evolutionary model have been inputting their philosophical assumptions into its construction and throughout its development, it might be that a panpsychist-informed scientific approach to evolution can bear greater fruit than physicalism. For example, perhaps panpsychism's apparent immunity to ontological concerns regarding the evolution of consciousness is a result of panpsychist motivations being upheld within its development and structure. Or, more strongly, perhaps panpsychism appears immune to such concerns because it is the only way of reconciling consciousness with evolution, and the developers of evolution knew this.⁸⁸ Now, this is not to assert the truth or falsity of physicalism,

⁸⁸ As before, this is merely indicative of future potential research, as opposed to presenting a specific argument in-and-of-itself.

biopsychism, or panpsychism, but merely intends to provide context for *why* physicalism – as traditionally defined – might struggle to give an evolutionary account of consciousness. At the very least, if panpsychism is a sentiment shared by some of the dominant formulators of our evolutionary model, then we would probably do well to take heed of it.

Interestingly, the biopsychist sentiment of biologists and evolutionists continues to this day. For example, Calvo et al. (2020) have recently proposed that there is a real biological benefit to regarding plants as intelligent both from the fundamental issue of understanding plant life, but also from providing a direction for fundamental future research and in crop breeding. Trewavas et al. (2019) has proposed that in order to retain parsimony with evolution, a basic form of consciousness must exist in the evolution of first cells and as an inherent feature of cellular life. Gagliano et al. (2016), has proposed that associative learning is an essential component of plant behaviour (and which requires consciousness as a prerequisite). Finally, Gershman et al. (2020), Bray (2009), Turner, (2002), and Jenuwein and Allis (2001) have discovered that single cells – such as paramecium – have a number of remarkable abilities, such as possessing a sophisticated form of memory, and possessing the ability to carry out a form of information processing that neuroscientists have typically reserved to large networks of cells.

While some of these scientific studies posit biopsychism so as to retain parsimony with the evolution of consciousness, others posit consciousness in all living things by virtue of their behaviour. In this manner, these studies suggest that we must posit consciousness within all living things not just to reconcile consciousness with evolution, but also because there are genuine behavioural reasons for their being classified as conscious rather than not. Accordingly, we can see that biologists and evolutionists continue to be drawn to biopsychism

for a number of reasons. However, while biopsychism itself might allow for the retention of a more consistent model of evolution, we have seen that it must also be grounded in the appropriate metaphysics if one's ontology is to be consistent. In the Section 5.3.3 we discussed the issues that biopsychist physicalism will run into and also considered the immunity of panpsychism to such worries. All things considered, it may then be useful in future to explore whether a panpsychist-informed scientific approach to evolution can bear fruit.

5.5 ~ Concluding Remarks

This chapter has proposed a new argument against physicalism which seeks to establish that Russellian panpsychism provides the most theoretically promising response to the Mind-Body Problem. To begin the chapter, I proposed that in order to establish the advantage of one theory over the other, progress must be made into explaining one of their respective problems away, or a new argument which worsens the problems of the other theory is needed. My conclusion in this chapter was achieved via this latter approach; I showed show physicalism runs into some damming issues with regards to the evolution of consciousness, whereas Russellian panpsychism does not. This chapter was also intended to serve two functions: first, to serve as a safeguard to Russellian panpsychism, such that even with the Combination Problem looming, Russellian panpsychism can still be taken to provide a more attractive theory of consciousness than physicalism; second, to provide a foundation for potential future responses to the Combination Problem to build from.

To establish that Russellian panpsychism offers the best available account of consciousness, I proposed my 'Argument from Evolution' against type-b and type-c physicalism. I began its presentation with a defence of two core claims: (1) that the physicalist should endorse the precision of consciousness, and (2) that they should endorse that consciousness resulted from

saltation, as opposed to being something fundamental. In the case of (1), I argued that the physicalist who maintains that consciousness is vague must still be able in principle to provide an explanation for what it is about those fundamental differences relevant to an entity's classification as either non-conscious, vaguely conscious, or conscious. However, I proposed that any answer will be necessarily similar in kind to that which is provided by the Russellian panpsychist, such that the former will risk reducing into the latter. While the physicalist could resist this claim by proposing vagueness in constitution – such that reference to categories of consciousness is mistaken – I showed how this creates a particularly dangerous Sorites paradox that either removes their credibility as a response to the Mind-Body Problem or again leads them to a panpsychist conclusion. So as to avoid these issues, the physicalist should endorse the precision of consciousness, and should also opt to provide their evolutionary account of consciousness via saltation (2), rather than proposing consciousness to be something fundamental.

Upholding these claims, and by working down our evolutionary trajectory with regards to our evolutionary relatives, I showed that the physicalist cannot find any non-arbitrary point where consciousness could have originated aside from the origin of living things, such that their thesis becomes 'biopsychist.' I documented that biopsychist physicalism will run into three dangerous problems that weaken physicalism further. The first is prevalent only to the type-c variant of physicalism: their Hard Problem has got insurmountably harder. By proposing a biopsychist account of physicalism, the type-c physicalist can no longer appeal to the abstract complexity of brains and bodies as the reason for the Hard Problems persistence. Instead, they must now uphold that the Hard Problem could be explained away through knowledge of the simplest living things. However, our understanding of the simplest living entities is relatively complete, and we can see nothing about their physical constitution which explains away the Hard

Problem. Indeed, while the explananda allegedly relevant to explaining it away have got simpler, the gap in our explanation regarding the explanandum has got larger.

The second issue that the biopsychist physicalist runs into is that both type-b and -c now possess a similar Combination Problem to Russellian panpsychism. Since humans are now composed of many ‘little’ conscious living things, the physicalist must give an account for how it is that they combine to make singular conscious living entity. Although the respective Combination Problems for type-b and -c possess no more dialectical force than that which is proposed against Russellian panpsychism, the biopsychist physicalist *also* retains their pervasive Hard Problem, such that they have all the problems of Russellian panpsychism, and more. With this in mind, the ‘level playing field’ between our two theses is no longer retained; Russellian panpsychism should now be theoretically preferred.

The final issue that both our physicalist positions run into, is that biopsychist physicalism might itself be an unstable position. I proposed that if ‘life’ is vague, it might not signify a non-arbitrary point for the origin of consciousness. If so, the biopsychist physicalist has three further options they can explore. First, they might choose to posit the origin of consciousness somewhere prior to evolution, such that it originates with something non-living and more fundamental than the most basic living things. However, this would dissolve any significant difference between their thesis and Russellian panpsychism (whilst still maintaining the Hard Problem). Second, the physicalist might instead just opt to tie consciousness to life, and so concede that consciousness, like life, must also be vague. However, doing so would cause them to fall victim to the aforementioned issues regarding consciousness as vague, and again force a potential reduction of physicalism into Russellian panpsychism. Finally, the physicalist could attempt to find some non-arbitrary distinction in the physical constitution of those things in

distinct categories of our now vague spectrum of life. They could then tie such a non-arbitrary point to its equivalent distinction in the spectrum of consciousness, and so provide a non-arbitrary point for the origin of consciousness that still endorses biopsychism. However, as noted, this would require that there is something about fundamental entities relevant to the instantiation of consciousness, and so, like before, the biopsychist physicalist will risk his doctrine reducing into something similar to Russellian panpsychism.

After setting up these problems for physicalism, I briefly considered the invulnerability of Russellian panpsychism to the Argument from Evolution. Since quiddities are the non-structural, non-relational, categorical properties which underlie and characterise the micro-level entities posited by physics, and since quiddities are phenomenal properties, phenomenal properties underlie and characterise the micro-level entities posited by physics in its characterisation of matter. In this manner, consciousness is the categorical property of fundamental entities, and so it resides within every composite thing in the precise sense, regardless of whether it is combined into a singular macro-subject. I argued that the Russellian panpsychist, therefore, doesn't have to give an account as to *when* consciousness pops up within our evolutionary trajectory, since consciousness comes before evolution. In this light, I concluded that panpsychism inherits no further issues from the Argument from Evolution over and above its pre-existing Combination Problem, but crucially, that same issue is now one of the many also suffered by physicalism.

The Argument from Evolution has threatened physicalism's reduction into Russellian monism, its inheritance of the Combination Problem, and the worsening of its Hard Problem. On the other hand, Russellian panpsychism has been shown to be invulnerable to the Argument from Evolution, and to possess fewer theoretical issues than physicalism. With this in mind, and all

other things being equal, Russellian panpsychism provides the more theoretically promising response to the Mind-Body Problem.

While the bulk of this chapter has argued that Russellian panpsychism offers the best available account of consciousness, I also considered some potential future ways in which the Russellian panpsychist could build from the Argument from Evolution in response to the Combination Problem. In this manner, while the Argument from Evolution provided Russellian panpsychism with a theoretical ‘safety net’ over physicalism with regards to establishing its internal coherence, it also provides it with the potential grounding to resolve its own internal issues in the future. I proposed that the Argument from Evolution presents three options for the Russellian panpsychist: the first, was to propose that combination occurs prior to evolution; the second, was to propose that combination occurs as a process of evolution; the third, was to propose that combination occurs in all living things, such that combination coincides with the origin of life. While my analyses into these options was intended solely to be indicative of future interesting avenues of research, I showed how each of these options could prove fruitful for the Russellian panpsychist.

I then briefly considered some potentially informative historical context for the conclusions of this chapter. In particular, I explored potential evidence that some of the dominant proponents and developers of the evolutionary model – namely, Charles Darwin, George Romanes, and Sewall Wright – believed in the biopsychist thesis, and discovered that while Darwin does not explicitly ground this thesis in any metaphysical account of consciousness, Romanes and Wright ground theirs in panpsychism. While such considerations are again only indicative of potential future research, I explained how the philosophical assumptions of these thinkers might have been built into their model of evolution. After briefly considering how this

biopsychist sentiment is still present within the literature on evolution today, I harnessed the previous conclusions of this chapter to suggest that a panpsychist-informed scientific approach to evolution might be a beneficial avenue for future research.

Conclusion

This thesis has proposed that Russellian panpsychism provides the best available solution to the Mind-Body Problem. Specifically, I proposed that even with the Combination Problem looming, Russellian panpsychism can still offer a better solution to the Mind-Body Problem than the dominant approaches of type-b and type-c physicalism. Although Chalmers' 2-Dimensional Conceivability Argument against physicalism is commonly used by Russellian panpsychists to help motivate this conclusion, I acknowledged how the Combination Problem can be used to create a similar Conceivability Argument against Russellian panpsychism. In this manner, I concluded that Conceivability Arguments are of no help to the Russellian panpsychist, as they are themselves vulnerable to it.

Through assessing the breadth of application of Conceivability Arguments, alongside how their consistent use renders contradictions, I proposed that rather than showing physicalism or Russellian panpsychism to be false, these Conceivability Arguments are themselves flawed. In particular, I showed how Conceivability Arguments can be levelled to *all* philosophy of mind hypotheses, such as Russellian panprotopsychism, substance dualism, and property dualism (Goff, 2017; Chalmers, 2016b; Aleksiev, 2023; Brown, 2010; Cutter, 2019). I then argued that – assuming each argument has equal force – there are three clear possibilities that one may draw. Either: (i) each of my proposed arguments' premises are true in each application, and that our entire spectrum of hypotheses are incorrect; or, (ii) at least one of their premises is false in each application, and so they cannot provide us with a true conclusion about any philosophy of mind hypothesis; or, (iii) an adequate justification for the truth of at least one of

their premises is currently unattainable, and so they are currently of no use within the Mind-Body debate.

To elucidate such matters further, I constructed a series of valid Conceivability Arguments which negate a metaphysical truth about the world from which they are posited. To begin, I introduced Katalin Balog criticism of Conceivability Arguments (1999), where she suggests that the conceived world within our initial 2-Dimensional Conceivability Argument is one in which physicalism is necessarily true. Crucially, Balog proposes that entities from within this world could produce an identical Conceivability Argument that will that negate the truth of physicalism. Because of this contradiction, and by virtue of the arguments being identical, Balog proposes that there must be a false premise within both arguments. However, I acknowledged a potential flaw in Balog's argument: her allegedly identical Conceivability Argument does not uphold the tacit premise 'someone is phenomenally conscious' that is present in the original. I acknowledged that without this tacit premise, Balog cannot establish her intended conclusion.

To bypass this concern, I proposed three 'Counterfactual Conceivability Arguments' which each upheld this required tacit premise. I then harnessed a similar set of contradictions to illustrate that there must be (at least) a flawed premise within Conceivability Arguments. This analysis led me to the conclusion that the vulnerability of our philosophy of mind hypotheses to Conceivability Arguments is not necessarily a fault of those hypotheses themselves, but rather a fault within the argument itself. In this manner, I discounted consequence (i); it is not the case that each of my proposed arguments contain solely true premises, and that our entire spectrum of hypotheses are incorrect.

I then located this fault in P.1 of the Conceivability Argument, specifically with regards to the claim that any of the previously stated scenarios are *ideally* conceivable. I proposed that we do not have sufficient reason in any instance concerning Q to make judgements about what would be known from ideal perspectives. In particular, I proposed that any attempt to deliberate on what can be ruled out a priori from an ideal perspective necessitates an extrapolation from our current a priori knowledge, otherwise the conceivability remains of the *prima facie* form. However, in practice, to achieve this ideal perspective is then just to harness current assumptions about particular epistemic gaps, assume they will remain once in possession of an ideal knowledge, and then to conclude that from that ideal perspective, such gaps will remain impassable. Assuming that ‘ideal conceivability’ does actually refer to something distinct and stronger than *prima facie* conceivability, at least in our current application, we clearly do not achieve it.

Following this logic, I proposed that the visible contradiction in my Counterfactual Conceivability Arguments likely originates from a flawed application of what is ideally conceivable, rather than there being an implicit flaw necessarily shared amongst a premise of all Conceivability Arguments. I proposed that consequence (iii), therefore, is the most appropriate to draw from my analysis; we simply do not have an adequate justification for the truth for P.1 in each application, and so Conceivability Arguments are currently of no use within the Mind-Body debate. While I conceded that Conceivability Arguments could be informative in future, I showed how this would only obtain iff we already knew whether the hypothesis in question was true or false. However, since at this point, we would know the truth in question, Conceivability Arguments would be of no further use.

I ultimately concluded that Conceivability Arguments should be relinquished from the Russellian panpsychist's repertoire, and indeed, from the Mind-Body debate entirely. While doing so removes one of the most popularly appealed to arguments for the theory, such a move is pertinent if the Russellian panpsychist wishes to make a compelling case for their theoretical superiority over physicalism. In light of this, I began to structure a new argument for Russellian panpsychism and against physicalism, concerning the evolution of consciousness. My argument defended with two core claims: (1) that the physicalist *should* endorse the precision of consciousness, and (2) that the physicalist who endorses the precision of consciousness will endorse that consciousness is a result of saltation, as opposed to being something fundamental. I showed how the physicalist has good reason to endorse (1) and that the denial of this may have drastic consequences for their theory. Similarly, I showed how the physicalist must endorse (2), or else admit no intelligible difference between their theory and the panpsychist implications that they hope to deter.

Upholding these claims, and by working down our evolutionary trajectory with regards to our evolutionary relatives, I showed how the physicalist cannot find any non-arbitrary point where consciousness could have originated aside from the origin of the first living things. However, I argued that through conceding this, the physicalist runs into three issues that substantially diminish its credibility: the first, was that the Hard Problem for type-c physicalism gets insurmountably harder; the second, was that both type-b and -c physicalism now possess a similar Combination Problem to Russellian panpsychism; and, the third, was that biopsychist physicalism might itself be unstable – as 'life' might be a vague concept – and so the theory may still reduce into something similar to Russellian panpsychism.

Whilst primarily directed against physicalism, I showed how my argument also serves as a safeguard for Russellian panpsychism. Specifically, I showed how the Russellian panpsychist is invulnerable to my Argument from Evolution: it inherits no further issues from the argument over that which it already suffered – the Combination Problem – and that same issue is now also suffered by both type-b and type-c forms of physicalism. Since physicalism was shown to maintain all its own issues, whilst inheriting the most dangerous argument against Russellian monism, physicalism was concluded to be less theoretically promising than Russellian monism. In this manner, my argument established that even with its Combination Problem looming, Russellian panpsychism still offers a more attractive theory of consciousness than physicalism. *ceteris paribus*, Russellian panpsychism offers the best available response to the Mind-Body Problem.

I then explored how potential and future responses to the Combination Problem could be built from the logic of my argument. I explored three potential avenues of response: (a) combination occurs prior to evolution, (b) combination occurred as a result of evolution, (c) combination coincided with the origin of life. While option (a) has already received some support within the extant literature, I explored how (b) and (c) could also prove to be fruitful avenues for future research that may help establish the internal coherency of Russellian panpsychism. Finally, I explored some potential context for the conclusions of my thesis. In particular, I explored potential evidence that some of the most dominant proponents of the evolutionary model might have believed in something similar to biopsychism, and some thought necessary to ground that thesis in panpsychism (and indeed, build that assumption into their model). As a result, I explored the claim that a panpsychist-informed scientific approach to evolution may be useful in the future. While this analysis was wholly speculative, like those prospective responses to

the Combination Problem, the contextualisation for my Argument from Evolution may represent a fruitful avenue for future research.

I conclude that Russellian panpsychism offers the best available response to the Mind-Body Problem. While the Conceivability Argument was shown to be toothless in achieving this conclusion, I have constructed a new argument that serves as a safeguard for the theory, such that even with its Combination Problem unresolved, it can still offer more theoretically promising response to the Mind-Body Problem than type-b and type-c forms of physicalism: physicalism – unlike Russellian panpsychism – cannot give a satisfactory account of why and how consciousness evolved. Furthermore, since future research into the evolution of consciousness seems likely to elucidate genuine avenues for Russellian panpsychism to resolve its Combination Problem, not only does Russellian panpsychism offer a superior potential resolution to the Mind-Body Problem than physicalism, but it may also simply be true.

Bibliography

- Aleksiev, D. (2023). 'Idealist Panpsychism and Spacetime Structure.' *Journal of the American Philosophical Association*. Vol. 10. No. 3.
- Alexander, S. (1920). *Space, Time, and Deity*. (New York: Macmillan).
- Alter, T. (2016). 'The Structure and Dynamics Argument Against Materialism'. *Noûs*. Vol. 50. No. 4.
- Alter, T. & Nagasawa, Y. (2015). 'What is Russellian Monism?' *Consciousness in the Physical World: Perspectives on Russellian Monism*. (New York: Oxford University Press).
- Ameta, S.; Arsène, S.; Foulon, S.; Saudemont, B.; Clifton, B.E.; Griffiths, A.D.; Nghe, P. (2021). 'Darwinian Properties and Their Trade-Offs in Autocatalytic RNA Reaction Networks.' *Nat. Commun.* Vol. 12.
- Antony, M. V. (2006). 'Vagueness and the Metaphysics of Consciousness,' *Philosophical Studies*. Vol. 128. No. 3.
- Aquinas, T. (1947). *Summa Theologiae*. the Fathers of the English Dominican Province (trans.). (New York: Benziger Brothers).
- Armstrong, D. M. (1997). *A World of States of Affairs*. (Cambridge: Cambridge University Press).
- Armstrong, D. M. (1999). *The Mind-Body Problem: An Opinionated Introduction*. (New York: Routledge).
- Bailey, A. (2007). 'The Unsoundness of Arguments from Conceivability.' <https://philpapers.org/rec/BAITUO>.
- Balog, K. (1999). 'Conceivability, Possibility, and the Mind-Body Problem.' *The Philosophical Review*. Vol. 108. No. 4.

- Barnes, J. (1982). *The Presocratic Philosophers: The Arguments of the Philosophers*. (London and New York: Routledge).
- Bays, C. (2009). 'Gliders in Cellular Automata.' *Encyclopedia of Complexity and Systems Science*. R. A. Meyers (ed.) (London: Springer-Verlag).
- Bedau, M. (1996). 'The nature of life.' In *The Philosophy of Artificial Life*. M. Boden (ed.). (Oxford: Oxford University Press).
- Bedau, M. (1998). 'Four puzzles about life.' *Artificial intelligence*. Vol. 4.
- Bedau, M., & Cleland, C. (eds.), 2010, *The Nature of Life: Classical and Contemporary Perspectives from Philosophy and Science*. (New York: Cambridge University Press).
- Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., ... Canaider, S. (2013). 'An estimation of the number of cells in the human body.' *Annals of Human Biology*. Vol. 40. No. 6.
- Bird, A. (2007a). *Nature's Metaphysics: Laws and Properties*. (Oxford: Oxford University Press).
- Bird, A. (2007b). 'The Regress of Pure Powers?' *Philosophical Quarterly*. Vol. 57.
- Blackburn, S. (1990). 'Filling in Space.' *Analysis*. Vol. 50.
- Blain, J.C. & Szostak, J.W. (2014). 'Progress Toward Synthetic Cells.' *Annu. Rev. Biochem.* Vol. 83.
- Block, N. (1980). 'Troubles with Functionalism.' *Readings in the Philosophy of Psychology*. Vol. 1. (Cambridge, MA: Harvard University Press).
- Block, N. & Stalnaker, R. (1999). 'Conceptual Analysis, Dualism, and the Explanatory Gap.' *Philosophical Review*. Vol. 108.
- Braddon-Mitchell, D., & Jackson, F. (2006). *Philosophy of mind and cognition: An introduction*. (Malden, MA: Wiley-Blackwell).

- Bray D. (2009). *Wetware: A Computer in Every Living Cell*. (New Haven: Yale University Press).
- Broad, C. D. (1925). *The Mind and Its Place in Nature*. (Abingdon: Routledge & Kegan Paul).
- Brown, R. (2010). 'Deprioritising the A Priori Arguments Against Physicalism.' *Journal of Consciousness Studies*. Vol. 17.
- Brüntrup G. (2016). 'Emergent Panpsychism,' *Panpsychism: Contemporary Perspectives*. G. Brüntrup & L. Jaskolla (eds.) (Oxford: Oxford University Press).
- Brüntrup G. & Jaskolla, L. (2016). *Panpsychism: Contemporary Perspectives*. (Oxford: Oxford University Press).
- Bruylants, G.; Bartik, K.; Reisse, J. (2010). 'Is It Useful to Have a Clear-Cut Definition of Life? On the Use of Fuzzy Logic in Prebiotic Chemistry.' *Orig. Life Evol. Biosph.* Vol. 40.
- Calvo, P., Gagliano, M., Souza, G.M., and Trewavas, A. (2020). 'Plants are intelligent, here's how.' *Annals of Botany*. Vol. 125.
- Campbell, K. (1970). *Body and Mind*. (London: Doubleday).
- Campbell, K. (1976). *Metaphysics: An Introduction*. (Encino, CA: Dickenson).
- Carruthers, G., & Shier, E. (2012). 'Dissolving the Hard Problem of Consciousness,' C04 online conference.
- Cerullo, M. A. (2015). 'The Problem with Phi: A Critique of Integrated Information Theory.' *PLoS Comput Biol*.
- Chalmers, D. (1995). 'Facing Up to the Problem of Consciousness.' *Journal of Consciousness Studies*. Vol. 2. No. 3.
- Chalmers, D. (1996). *The Conscious Mind*. (Oxford: Oxford University Press).

- Chalmers, D. (2002). 'Consciousness and Its Place in Nature.' *Philosophy of Mind: Classical and Contemporary Readings*. D. Chalmers (ed.). (Oxford: Oxford University Press).
- Chalmers, D. (2004). 'Epistemic Two-Dimensional Semantics.' *Philosophical Studies*. Vol. 118.
- Chalmers, D. (2006). 'Strong and Weak Emergence.' *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*. P. Davies & P. Clayton (eds.) (Oxford: Oxford University Press).
- Chalmers, D. (2010). *The Character of Consciousness*. (Oxford: Oxford University Press).
- Chalmers, D. (2012). *Constructing the World*. (Oxford: Oxford University Press).
- Chalmers, D. (2016a). 'Panpsychism and Panprotopsychism.' *Panpsychism: Contemporary Perspectives*. G. Brüntrup & L. Jaskolla (eds.) (Oxford: Oxford University Press).
- Chalmers, D. (2016b). 'The Combination Problem for Panpsychism.' *Panpsychism: Contemporary Perspectives*. G. Brüntrup & L. Jaskolla (eds.) (Oxford: Oxford University Press).
- Chambliss, B. (2018). 'The mind–body problem.' *WIREs Cogn Sci*. Vol. 9. No. 4.
- Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. (Massachusetts: MIT Press).
- Churchland, P. S. (1997). 'The Hornswoggle Problem. *Explaining Consciousness: The Hard Problem*. J. Shear (ed.). (Massachusetts: MIT Press).
- Churchland, P. S. (2013). *Touching a Nerve*. (New York: W.W. Norton and Company).
- Churchland, P. M. (1984). *Matter and Consciousness*. (Massachusetts: MIT Press).

- Clifford, W. (1874). 'Body and Mind,' *Lectures and Essays*. (Cambridge: Cambridge University Press).
- Clifford, W. (1878). 'On the Nature of Things-in-Themselves,' *Mind*, Vol. 3. No. 9.
- Coleman, S. (2006). 'Being Realistic: Why Physicalism May Entail Panexperientialism,' *Journal of Consciousness Studies*. Vol. 13.
- Coleman, S. (2009). 'Mind under Matter.' *Mind That Abides*. D. Skrbina (ed.) (Amsterdam: Benjamins).
- Coleman, S. (2012). 'Mental Chemistry: Combination for Panpsychists.' *Dialectica*. Vol. 66.
- Coleman, S. (2015). 'Neurocosmology.' *Phenomenal Qualities: Sense, Perception, and Consciousness*. P. Coates & S. Coleman (eds.). (Oxford: Oxford University Press).
- Coleman, S. (2016). 'Panpsychism and Neutral Monism: How to Make Up One's Mind.' *Panpsychism: Contemporary Perspectives*. G. Brüntrup & L. Jaskolla (eds.) (Oxford: Oxford University Press).
- Cottingham, J. (1992). *The Cambridge Companion to Descartes*. (Cambridge: Cambridge University Press).
- Crane, T. (1999). 'The Mind-Body Problem.' *MIT Encyclopaedia of the Cognitive Sciences*. R. A. Wilson, F. C. Keil (eds.). (Cambridge, USA: MIT Press).
- Cutter, B. (2018). 'Paradise Regained: A Non-Reductive Realist Account of the Sensible Qualities.' *Australasian Journal of Philosophy*. Vol. 96. No. 1.
- Cutter, B. (2019). 'Against the Middle Ground: Why Russellian Monism is Unstable.' *Analytic Philosophy*. Vol. 60. No. 2.
- Dainton, B. (2011). "Review of Consciousness and Its Place in Nature." *Philosophy and Phenomenological Research*. Vol. 83. No. 1.

- Darwin, C. R. (1859). *The Origin of Species by Means of Natural Selection*. (Penguin Books: Harmondsworth).
- Darwin, C. R. (1880). *Movements and Habits of Climbing Plants*. (London: John Murray).
- Darwin, C. R. (1901). *The Descent of Man*. (London: John Murray).
- Darwin, C. R. (1903). *More Letters of Charles Darwin: A record of his work in a series of hitherto unpublished letters*. Vol. 2. F. Darwin (ed.) (London: John Murray).
- Darwin, C. R. (1975). *Natural Selection*. R. C. Stauffer (ed.) (Cambridge: Cambridge University Press).
- Darwin, C. R. (c. 1838). 'Old and Useless Notes,' *Darwin on Man: A Psychological Study of Scientific Creativity*. P. H. Barrett, H. Gruber (eds.). (1974). (Wildwood House: London).
- Dasgupta, S. (2014). 'The Possibility of Physicalism.' *Journal of Philosophy*. Vol 111. No. 9.
- Davies, M. K. & Humberstone, I. L. (1980). 'Two notions of necessity.' *Philosophical Studies*. Vol. 38.
- Deacon, T. (2012). *Incomplete Nature: How Mind Emerged from Matter*. (New York: W.W. Norton & Company).
- Delafield-Butt, J. (2007). 'Towards A Process Ontology of Organism: Explaining the Behaviour of a Cell.' *Applied Process Thought: Frontiers of Theory and Research*. T. Kelly & M. Dibben (eds.). (Paris: Ontos Verlag).
- Dennett, D. (1991). *Consciousness Explained*. (Boston: Little, Brown).
- Dennett, D. (2005). *Sweet Dreams*. (Massachusetts: MIT Press).
- DeRose, K. (1995). 'Solving the Skeptical Problem.' *Journal of Philosophy*. Vol. 92. No. 5.

- Descartes, R. (1985a). *Principles of Philosophy*. In J. Cottingham, R. Stoothoff, & D. Murdoch (trans.). *The Philosophical Writings of Descartes: Volume I* (Cambridge: Cambridge University Press).
- Descartes, R. (1985b). *Meditations on First Philosophy*. In J. Cottingham, R. Stoothoff, & D. Murdoch (trans.). *The Philosophical Writings of Descartes: Volume II* (Cambridge: Cambridge University Press).
- Descartes, R. (1985c). *Passions of the Soul*. In J. Cottingham, R. Stoothoff, & D. Murdoch (trans.). *The Philosophical Writings of Descartes: Volume I*. (Cambridge: Cambridge University Press).
- Di Paolo, E.A. (2018). 'The enactive conception of life.' *The Oxford Handbook of 4E Cognition*. Newen, A., De Bruin, L. & Gallagher, S. (eds.). (Oxford: Oxford University Press).
- Di Paolo, E.A., Buhrmann, T. & Barandiaran, X. (2017). *Sensorimotor Life: An Enactive Proposal*. (Cambridge, MA: MIT Press).
- Di Paolo, E.A., Cuffari, E. & De Jaegher, H. (2018). *Linguistic Bodies: The Continuity Between Life and Language*. (Cambridge, MA: MIT Press).
- Dretske, F. (1995). *Naturalizing the Mind*. (Massachusetts: MIT Press).
- Eddington, A. (1920). *Space, Time and Gravitation: An Outline of the General Relativity Theory*. (Cambridge: Cambridge University Press).
- Eddington, A. (1927). *The Nature of the Physical World: Gifford Lectures of 1927*. (Cambridge: Cambridge Scholars Publishing).
- Ehret, G., Romand, R. (2022). 'Awareness and consciousness in humans and animals – neural and behavioural correlates in an evolutionary perspective.' *Frontiers in Systems Neuroscience*. Vol. 16.
- Eklund, M. (2005). 'What vagueness consists in.' *Philosophical Studies*. Vol. 125.

- Ellis, B. (2001). *Scientific Essentialism*. (Cambridge: Cambridge University Press).
- Ellis, B. (2002). *The Philosophy of Nature: A Guide to the New Essentialism*. (Ontario/Montreal: McGill-Queen's University Press).
- Evans, G. (1979). 'Reference and Contingency.' *The Monist*. Vol. 62.
- Feigl, H. (1960). 'Mind-Body, not a Pseudo-Problem.' *Dimensions of Mind*. Sidney Hook (ed). (New York: Collier Books).
- Feigl, H. (1962). 'The "Mental" and the "Physical"' *Minnesota Studies in the Philosophy of Science*. E. Lepore & D. Schramm (eds.). (Minnesota: University of Minnesota Press).
- Field, H. (1994). 'Disquotational Truth and Factually Defective Discourse,' *Philosophical Review*. Vol. 103.
- Fine, K. (2001). 'The Question of Realism,' in *Metaphysics, Vol 15: Philosophical Perspectives*. (New Jersey: Wiley Blackwell).
- Fine, K. (2012). 'A Guide to Ground,' in *Metaphysical Grounding*. F. Correia & B. Schneider (eds.). (Cambridge: Cambridge University Press).
- Fleming, S., Frith, C. D., Goodale, M., Lau, H., LeDoux, J. E., Lee, A. L. F., Michel, M., Owen, A. M., Peters, M. A. K. Slagter, H. A. (2015). 'The Integrated Information Theory of Consciousness as Pseudoscience.' <https://doi.org/10.31234/osf.io/zsr78> (Accessed 06/12/2023, 14:32).
- Fodor, J. A. (1975). *The Language of Thought*. (Massachusetts: Harvard University Press).
- French, S. (2013). *The Structure of the World*. (Oxford: Oxford University Press).
- Gagliano, M., Vyazovskiy, V., Borbély, A. (2016). 'Learning by Association in Plants.' *Sci Rep*. Vol. 6.

- Galileo, G. (1623). 'The Assayer.' 1, *Discoveries and Opinions of Galileo*. S. Drake (ed.) (New York: Anchor Books).
- Gershman, S.J., Balbi, P.E.M., Gallistel, C.R., Gunawardena, J. (2020). 'Reconsidering the evidence for learning in single cells.' *Computational and Systems Biology*.
- Gibb, S. (2015). 'Defending Dualism.' *Proceedings of the Aristotelian Society*. Vol. 115.
- Ginsburg, S. & Jablonka, E. (2020). 'Are your cells conscious?' *American Journal of Psychology*. Vol. 133. No. 1.
- Goff, P. (2006). 'Experiences Don't Sum.' *Journal of Consciousness Studies*. Vol. 13.
- Goff, P. (2009). 'Why Panpsychism Doesn't Help Us Explain Consciousness,' *Dialectica*. Vol. 63. No. 3.
- Goff, P. (2014). 'Orthodox Property Dualism + The Linguistic Theory of Vagueness = Panpsychism' *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Structure of Experience*. R. Brown (ed.). (New York: Springer).
- Goff, P. (2016). 'The Phenomenal Bonding Solution to the Combination Problem.' *Panpsychism: Contemporary Perspectives*. G. Brüntrup & L. Jaskolla (eds.). (Oxford: Oxford University Press).
- Goff, P. (2017). *Consciousness and Fundamental Reality*. (Oxford: Oxford University Press).
- Goff, P. (2019). *Galileo's Error: Foundations for a New Science of Consciousness*. (London: Rider).
- Goff, P. & Coleman, S. (2020). 'Russellian Monism,' *The Oxford Handbook of the Philosophy of Consciousness*. U. Kriegel (ed.). (Oxford: Oxford University Press).

- Gould, S. (1994). 'The Evolution of Life on the Earth.' *Scientific American*. Vol. 241. No. 4.
- Greenough, P. (2003). 'Vagueness: A minimal theory.' *Mind*. Vol. 112. No. 446.
- Gunz, P., Neubauer, S., Maureille, B., Hublin, J-J. (2010). 'Brain development after birth differs between Neanderthals and modern humans.' *Current Biology*. Vol. 20. Issue: 21.
- Haeckel, E. (1892). 'Our monism: The principles of a consistent, unitary world-view.' *The Monist*. Vol. 2.
- Hall, G. (2023). 'Is Consciousness Vague?' *Australasian Journal of Philosophy*. Vol. 101. No. 3.
- Hameroff, S. & Penrose, R. (1996). 'Conscious Events as Orchestrated Space-Time Selections.' *Journal of Consciousness Studies*. Vol. 3. No. 1.
- Harman, G. (1990). 'The Intrinsic Quality of Experience.' *Philosophical Perspectives*. Vol. 4.
- Hawking, S. (1992). *A Brief History of Time: From the Big Bang to Black Holes*. (Bantam Press: London)
- Hawthorne, J. (2001). 'Causal Structuralism.' *Philosophical Perspectives*. Vol. 15.
- Heil, J. (2003). *From an Ontological Point of View*. (Oxford: Clarendon Press).
- Heller, M. (1988). 'Vagueness and the Standard Ontology.' *Nous*. Vol. 22. No. 1.
- Herculano-Houzel, S. (2009). 'The human brain in numbers: a linearly scaled-up primate brain.' *Front Hum Neurosci*. Vol. 9. No. 3.
- Herculano-Houzel, S. (2013). 'The Remarkable, Yet Not Extraordinary, Human Brain as a Scaled-Up Primate Brain and Its Associated Cost.' *In the Light of Evolution: Volume VI: Brain and Behaviour*. G. F. Striedter, J. C. Avise, F. J. Ayala (eds). (Washington: The National Academies Press).

- Herculano-Houzel, S., Kaas, J. (2011). 'Gorilla and Orangutan Brains conform to the Primate Cellular Scaling Rules: Implication for Human Evolution.' *Brain Behav Evol.* Vol. 77. Issue 1.
- Hill, C. S. (1997). 'Imaginability, Conceivability, Possibility, and the Mind-body Problem.' *Philosophical Studies.* Vol. 87.
- Hodge, M.J.S. (1992). 'Biology and Philosophy (Including Ideology): A Study of Fisher and Wright,' *Boston Studies in the Philosophy of Science.* Vol. 142.
- Hofman M. (1985). 'Size and shape of the cerebral cortex in mammals. I. The cortical surface.' *Brain Behav Evol.* Vol. 27.
- Holman, E. L. (2008). 'Panpsychism, Physicalism, Neutral Monism and the Russellian Theory of Mind.' *Journal of Consciousness Studies.* Vol. 15.
- Horgan, T. (1994). 'Robust Vagueness and the Forced-March Sorites Paradox,' *Philosophical Perspectives.* Vol. 8.
- Horgan, T. (2010). 'Transvaluationism About Vagueness: A Progress Report.' *The Southern Journal of Philosophy.* Vol. 48. No. 1.
- Horowitz, N.H., Miller, S.L. (1962). Current Theories on the Origin of Life. *Progress in the Chemistry of Organic Natural Products.* Zechmeister, L., (ed). (Springer: Vienna, Austria)
- Jackson, F. (1982). "Epiphenomenal Qualia." *Philosophical Quarterly.* Vol. 32. No. 127.
- Jackson, F. (1998). *From Metaphysics to Ethics.* (New York: Oxford University Press).
- James, W. (1890). *The Principles of Psychology.* (New York: Dover).
- James, W. (1909). *A Pluralistic Universe: Hibbert Lectures to Manchester College on the Present Situation in Philosophy.* (London: Longmans, Green).
- Jenuwein T., Allis, C.D. (2001). 'Translating the histone code.' *Science.* Vol. 293

- Joyce, G.F. (1994). Foreword. *In Origins of Life: The Central Concepts*. Deamer, D.W., Fleischaker, G.R., (eds). (Oxford: Jones and Barlett)
- Kaas, J. (2013). 'The Evolution of Brains from Early Mammals to Humans.' *Cognitive Science*. Vol. 4. Issue 1.
- Kim, J. (1993). *Supervenience and Mind: Selected Philosophical Essays*. (Cambridge: Cambridge University Press)
- Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. (Massachusetts: MIT Press).
- Kim, J. (2005). *Physicalism, or Something Near Enough*. (New Jersey: Princeton University Press).
- Kim, J. (2010). *Philosophy of mind*. (Boulder, CO: Westview Press).
- Kirby, W. (1835). *On the Power, Wisdom, and Goodness of God, as Manifested in the Creation of Animals and in their History, Habits and Instincts*. (London: William Pickering).
- Kirk, R. (1974). 'Zombies vs. Materialists.' *Proceedings of the Aristotelian Society*. Vol. 48.
- Koch, C. (2014). 'Ubiquitous Minds.' *Scientific American Mind*. Vol. 25. No. 1.
- Koch, C. & Tononi, G. (2015). 'Consciousness: Here, There, and Everywhere?' *Philosophical Transactions B*. Vol. 370.
- Kripke, S. A. (1980). *Naming and Necessity*. (Cambridge, MA.: Harvard University Press).
- La Scola, B.; Audic, S.; Robert, C.; Jungang, L.; de Lamballerie, X.; Drancourt, M.; Birtles, R.; Claverie, J.-M.; Raoult, D. A. (2003). 'Giant Virus in Amoebae.' *Science*. Vol. 299.

- Ladyman, J. (1998). 'What Is Structural Realism?' *Studies In History and Philosophy of Science*. Vol. 29.
- Ladyman, J., Ross, D., Spurrett, D., & Collier, J. (2007). *Everything Must Go: Metaphysics Naturalised*. (Oxford: Oxford University Press).
- Lee, D. (forthcoming). 'Reinterpreting Anne Conway: A Cosmopsychist Ontology of Created Substance?' *Sophia*.
- Leibniz, G. W. (1898). *Monadology and Other Philosophical Writings*. R. Latta (trans.). (Oxford: Oxford University Press).
- Leibniz, G. W. (1989). *Philosophical Essays*. 'Mondadology.' R. Ariew and D. Garber (Trans.) (Indianapolis: Hackett).
- Levin, J. (2012). 'Do Conceivability Arguments Against Physicalism Beg the Question?' *Philosophical Topics*. Vol. 40. No. 2.
- Levine, J. (1983). 'Materialism and Qualia: The Explanatory Gap.' *Pacific Philosophical Quarterly*. Vol. 64.
- Lewis, D. (1986). *On the Plurality of Worlds*. (Oxford: Blackwell).
- Lewis, D. (1988). 'What Experience Teaches.' *Proceedings of the Russellian Society*. (Sydney: University of Sydney Press).
- Lewis, D. (1997). 'Finkish Dispositions.' *The Philosophical Quarterly*. Vol. 47. No. 187.
- Loar, B. (1990). 'Phenomenal States.' *Philosophical Perspectives*. Vol. 41.
- Lowe, E. J. (1998). *The Possibility of Metaphysics*. (Oxford, Oxford University Press).
- Lowe, E. J. (2000). *An Introduction to the Philosophy of Mind*. (Cambridge: Cambridge University Press).
- Lowe, E. J. (2006). *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. (Oxford: Oxford University Press).

- Lucretius. (1910). *On the Nature of Things*. Book 2. Cyril Bailey (trans.). (Oxford: Clarendon Press).
- Lycan, W. G. (1996). *Consciousness and Experience*. (Massachusetts: MIT Press).
- Malaterre, C. (2024). 'Is Life Binary or Gradual?' *Life*. Vol. 14.
- Mallozzi, A. (2021). 'Two Notions of Metaphysical Modality.' *Synthese*. Vol. 198. (S6).
- Marcus, E. (2004). 'Why Zombies Are Inconceivable.' *Australasian Journal of Philosophy*. Vol. 82.
- Marino, L., Reiss, D., Gallup, GG Jr. (1994). 'Mirror self-recognition in bottlenose dolphins: implications for comparative investigations of highly dissimilar species.' Parker ST, Mitchell RW, Boccia ML (eds). *Self-awareness in animals and humans: developmental perspectives*. (Cambridge: Cambridge University Press).
- Marten, K. Psarakos, S. (1994). 'Evidence of self-awareness in the bottlenose dolphin (*Tursiops truncatus*).' Parker ST, Mitchell RW, Boccia ML (eds) *Self-awareness in animals and humans: developmental perspectives*. (Cambridge: Cambridge University Press).
- Marten, K. Psarakos, S. (1995a). 'Marten and Psarakos commentary response.' *Conscious Cognition*. Vol. 4.
- Marten, K. Psarakos, S. (1995b). 'Using self-view television to distinguish between self-examination and social behavior in the bottlenose dolphin (*Tursiops truncatus*).' *Conscious Cognition*. Vol. 4
- McClelland, T. (2013). 'The Neo-Russellian Ignorance Hypothesis: A Hybrid Account of Phenomenal Consciousness.' *Journal of Consciousness Studies*. Vol. 20. No. 3–4.
- McGinn, C. (1989). 'Can We Solve the Mind-body Problem?' *Mind*. Vol. 98.
- McGinn, C. (1996). *The Character of Mind*. (2nd ed.). (Oxford: Oxford University Press).

- Melnyk, A. (2003). *A Physicalist Manifesto: Thoroughly Modern Materialism*. (Cambridge: Cambridge University Press).
- Merricks, T. (2001). 'Varieties of Vagueness.' *Philosophy and Phenomenological Research*. Vol. 62. No. 1.
- Miller, G. (2018). *The Combination Problem for Panpsychism: A Constitutive Russellian Solution*. Ph.D. thesis, University of Liverpool.
- Miller, J. (1978). *The Body in Question*. (New York: Random House).
- Mizrahi, M. & Morrow, D. (2014). 'Does Conceivability Entail Metaphysical Possibility?' *Ratio*. Vol. 28. No. 1.
- Molnar, G. (2003). *Powers: A Study in Metaphysics* (Oxford: Oxford University Press).
- Montero, B. (2013). 'Must Physicalism Imply the Supervenience of the Mental on the Physical?' *The Journal of Philosophy*. Vol. 110. No. 2.
- Montero, B. (2015). 'Russellian Physicalism.' *Consciousness in the Physical World: Perspectives on Russellian Monism*. T. Alter & Y. Nagasawa (eds.) (New York: Oxford University Press).
- Morange, M. (2012). 'The Recent Evolution of the Question "What is life?"'. *History and Philosophy of the Life Sciences*. Vol. 34. No. 3.
- Mørch, H. H. (2014). *Panpsychism and Causation: A New Argument and a Solution to the Combination Problem*, Ph.D. thesis, University of Oslo.
- Mumford, S. (1998). *Dispositions*. (Oxford: Oxford University Press).
- Mumford, S. (2004). *Laws in Nature* (London: Routledge).
- Nagasawa, Y. & Wager, K. (2016). 'Panpsychism and Priority Cosmopsychism.' *Panpsychism: Contemporary Perspectives*. G. Brüntrup & L. Jaskolla (eds.) (Oxford: Oxford University Press).
- Nagel, T. (1974). 'What Is It Like to Be a Bat.' *Philosophical Review*. Vol. 83.

- Nagel, T. (1986). *The View from Nowhere*. (Oxford: Oxford University Press).
- Nagel, T. (2007). 'What is the Mind-Body Problem.' *Experimental and Theoretical Studies of Consciousness: No. 74*. G. R. Bock, J. Marsh (eds.) (New Jersey: Wiley-Blackwell).
- Nakabachi, A.; Yamashita, A.; Toh, H.; Ishikawa, H.; Dunbar, H.E.; Moran, N.A.; Hattori, M. (2006). 'The 160-Kilobase Genome of the Bacterial Endosymbiont *Carsonella*.' *Science*. Vol. 314.
- Ney, A. (2008). 'Defining Physicalism,' *Philosophy Compass*. Vol 3. No. 5.
- Ney, A. (2015). 'A Physicalist Critique of Russellian Monism.' *Consciousness in the Physical World: Perspectives on Russellian Monism*. T. Alter & Y. Nagasawa (eds.) (New York: Oxford University Press).
- O'Rourke, J. (2018). *Property Dualism as a Solution to the Mind-Body Problem*. PhD Thesis, Princeton.
- Otto, S. (2022). 'An Approach to the De Novo Synthesis of Life.' *Acc. Chem. Res.* Vol. 55.
- Pályi, G., Zucchi, C., Caglioti, L. (2002) *Fundamentals of Life*. (Paris: Elsevier)
- Papineau, D. (1993). *Philosophical Naturalism*. (Oxford: Blackwell).
- Papineau, D. (2001). "The Rise of Physicalism." *Physicalism and Its Discontents*. In C. Gillett & B. Loewer (eds.). (Cambridge: Cambridge University Press).
- Papineau, D. (2002). *Thinking about Consciousness*. (Oxford: Oxford University Press).
- Penrose, R. (1996). *Shadows of the Mind: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*. (Oxford: Oxford University Press).
- Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*. (New York: Oxford University Press).

- Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. (Massachusetts: MIT Press).
- Planck, M. (1931). 'Interviews with Great Scientists.' *The Observer*. January 25th.
- Popa, R. (2004) *Between Necessity and Probability: Searching for the Definition and Origin of Life*. (New York: Springer)
- Princess Elisabeth of Bohemia and Descartes. (2007). *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*. L. Shapiro (Trans.) (Chicago: University of Chicago Press)
- Prinz, J. J. (2011). 'The Sensory Basis of Cognitive Phenomenology' in *Cognitive Phenomenology*. T. Bayne & M. Montague (eds). (Oxford: Oxford Academic).
- Prinz, J. J. (2012). *The Conscious Brain*. (Oxford: Oxford University Press).
- Provine, W. (1986). *Sewall Wright and Evolutionary Biology*. (Chicago: University of Chicago Press).
- Putnam, H. (1967). 'Psychological Predicates.' In *Art, Mind and Religion*. W. H. Capitan & D. D. Merrill (eds.). (Minnesota: University of Minnesota Press).
- Putnam, H. (1973). 'Meaning and Reference.' *Journal of Philosophy*. Vol. 70.
- Putnam, H. (1975). 'The Meaning of "Meaning"' *Mind, Language, and Reality*. (Cambridge: Cambridge University Press).
- Raatikainen, P. (2005). 'On the Philosophical Relevance of Gödel's Incompleteness Theorems. *Internationale de Philosophie*. Vol. 59.
- Reber, A.S. (2019) *The First Minds: Caterpillars, 'Karyotes, and Consciousness*. (New York: Oxford University Press).
- Reber, A.S. & Baluška, F. (2021). 'Cognition in some surprising places,' *Bio- chemical and Biophysical Research Communications*. Vol. 564.

- Rey, G. (1995). 'Toward a Projectivist Account of Conscious Experience.' *Conscious Experience*. T. Metzinger (ed.). (Kansas: Allen).
- Robinson, H. (1982). *Matter and Sense: A Critique of Contemporary Materialism*. (Cambridge: Cambridge University Press).
- Robinson, H. (1993). *Objections to Physicalism*. (New York: Oxford University Press).
- Robinson, H. (2004). *Mind and Consciousness: A Philosophical Introduction*. (Abingdon: Routledge).
- Robinson, H. (2016). *From the Knowledge Argument to Mental Substance*. (Cambridge: Cambridge University Press).
- Roelofs, L. (2019). *Combining Minds: How to Think About Composite Subjectivity*. (Oxford: Oxford University Press).
- Roelofs, L. (2021). 'Is Panpsychism at Odds with Science?' *Journal of Consciousness Studies*. Vol. 28.
- Romanes, G. J. (1882). 'The Fallacy of Materialism.' *The Nineteenth Century*. Vol. 12.
- Romanes, G. J. (1883). *Mental Evolution in Animals*. (London: John Murray).
- Rorty, R. (1979). *Philosophy and the Mirror of Nature*. (New Jersey: Princeton University Press).
- Rosen, G. (2010). 'Metaphysical Dependence: Grounding and Reduction.' *Modality: Metaphysics, Logic, and Epistemology*. B. Hale & A. Hoffman (eds.). (Oxford University Press).
- Rosenberg, G. (2004). *A Place for Consciousness*. (New York: Oxford University Press).
- Ruse, M. (2004). *Darwinian Heresies*. (Cambridge: Cambridge University Press).
- Russell, B. (1927a). *The Analysis of Matter*. (Nottingham: Spokesman).

- Russell, B. (1927b). *An Outline of Philosophy*. (London: Routledge).
- Russell, B. (1948). *Human Knowledge: Its Scope and Limits*. (London: Routledge).
- Russell, B. (1959). *My Philosophical Development*, (London: Routledge, 1995 edition).
- Russell, B. (1984). *Theory of Knowledge: The 1913 Manuscript*. (London: Routledge).
- Ryle, G. (1949). *The Concept of Mind*. (London: Hutchinson).
- Sarnat, H., Netsky, M. (2002). 'When Does a Ganglion Become a Brain? Evolutionary Origin of the Central Nervous System.' *Seminars in Pediatric Neurology*. Vol. 9. No. 4.
- Schaffer, J. (2010). 'Monism: The Priority of the Whole,' *Philosophical Review*. Vol. 119.
- Schiffer, S. (2003). *The Things We Mean*. (Oxford: Oxford University Press).
- Schrödinger, E. (1931). 'Interviews with Great Scientists.' *The Observer*. January 11th.
- Schroer, R. (2010). 'Is There More Than One Categorical Property?' *The Philosophical Quarterly*. Vol. 60. No. 241.
- Schwitzgebel, E. (2023). 'Borderline consciousness, when it's neither determinately true nor determinately false that experience is present.' *Philosophical Studies*. Vol. 180.
- Seager, W. (1995). 'Consciousness, Information, and Panpsychism.' *Journal of Consciousness Studies*. Vol. 2. No. 3.
- Seager, W. (2006). 'The 'Intrinsic Nature' Argument for Panpsychism.' *Journal of Consciousness Studies*. Vol. 13. No. 10–11.
- Seager, W. (2016). 'Panpsychism Infusion.' *Panpsychism: Contemporary Perspectives*. G. Brüntrup & L. Jaskolla (eds.). (Oxford: Oxford University Press).
- Shani, I. (2015). 'Cosmopsychism: A Holistic Approach to the Metaphysics of Experience.' *Philosophical Papers*. Vol. 44. No. 3.
- Shapiro, L. (2007). *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*. (Chicago: University of Chicago Press).

- Sider, T. (2001). *Four-Dimensionalism: An Ontology of Persistence and Time*. (Oxford: Oxford University Press).
- Sider, T. (2013). *Writing the Book of the World*. (Oxford: Oxford University Press).
- Simon, J. (2017). 'Vagueness and Zombies. Why 'Phenomenally Conscious has no Borderline Cases'.' *Philosophical Studies*.
- Skrbina, D. (2005). *Panpsychism in the West*. (Cambridge: MIT Press).
- Smart, J. J. C. (1959). 'Sensations and Brain Processes.' *The Philosophical Review*. Vol. 68. No. 2.
- Smith, C. U. M. (1978). 'Charles Darwin, the Origin of Consciousness, and Panpsychism.' *Journal of the History of Biology*, Vol. 11. No. 2.
- Smith, H. H.; Hyde, A. S.; Simkus, D. N.; Libby, E.; Maurer, S. E.; Graham, H. V.; Kempes, C. P.; Sherwood Lollar, B.; Chou, L.; Ellington, A. D.; et al. (2021) 'The Grayness of the Origin of Life.' *Life*. Vol. 11.
- Smith, N. J. J. (2005). 'Vagueness as closeness.' *Australasian Journal of Philosophy*. Vol 83. No. 2.
- Soames, S. (1999). *Understanding Truth*. (New York: Oxford University Press).
- Soames, S. (2004). *Reference and Description*. (Princeton, NJ: Princeton University Press).
- Sorensen, R. (1988). *Blindspots*. (Oxford: Clarendon Press).
- Steffes, D. (2006). 'Panpsychic Organicism: Sewall Wright's Philosophy for Understanding Complex Genetic Systems.' *Journal of the History of Biology*. Vol. 40.
- Stoljar, D. (2001). 'Two Conceptions of the Physical,' *Philosophy and Phenomenological Research*. Vol. 62. No. 2.
- Stoljar, D. (2006). *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*. (Oxford: Oxford University Press).

- Stoljar, D. (2010). *Physicalism*. (Abingdon: Routledge).
- Stoljar, D. (2015). 'Russellian Monism or Nagelian Monism?' *Consciousness in the Physical World: Perspectives on Russellian Monism*. T. Alter & Y. Nagasawa (eds.) (New York: Oxford University Press).
- Strawson, G. (1986). *Freedom and Belief*. (Oxford: Clarendon Press).
- Strawson, G. (1994). *Mental Reality*. (Cambridge, MA: MIT Press).
- Strawson, G. (2008). 'Realistic Monism: Why Physicalism Entails Panpsychism.' *Real Materialism and Other Essays*. Galen Strawson (ed.) (Oxford: Oxford University Press).
- Strawson, G. (2015). 'Real Materialism.' *Consciousness in the Physical World: Perspectives on Russellian Monism*. T. Alter & Y. Nagasawa (eds.) (New York: Oxford University Press).
- Strawson, P. (1964). *Individuals: An Essay in Descriptive Metaphysics*. (Abingdon: Routledge)
- Stubenberg, L. (2015). 'Russell, Russellian monism, and Panpsychism.' *Consciousness in the Physical World: Perspectives on Russellian Monism*. T. Alter & Y. Nagasawa (eds.) (New York: Oxford University Press).
- Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*, (Cambridge, MA: Harvard University Press).
- Thompson, E. (2022). 'Could All Life Be Sentient?' *Journal of Consciousness Studies*. Vol. 29. No. 3-4.
- Tononi, G. (2012). 'Integrated Information Theory of Consciousness: An Updated account.' *Archives Italiennes de Biologie*. Vol. 150. No. 2-3.
- Trewavas, A., Baluška, F., Mancuso, F., Calvo, P. (2019). 'Consciousness Facilitates Plant Behavior.' *Trends in Plant Science*. Vol. 25. No. 3.

- Trogon, K. (2013). 'Grounding: Necessary or Contingent?' *Pacific Philosophical Quarterly*. Vol. 94. No. 4.
- Turner, B. M. (2002). 'Cellular memory and the histone code.' *Cell*. Vol. 111.
- Tye, M. (1990). "Vague Objects," *Mind*. Vol. XCIX. No. 393
- Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. (Massachusetts: MIT Press).
- Tye, M. (2000). *Consciousness, Color, and Content*. (Massachusetts: MIT Press).
- Tye, M. (2021). *Vagueness and the Evolution of Consciousness: Through the Looking Glass* (Oxford: Oxford Academic).
- Tye, M. & Wright, B. (2011). 'Is There a Phenomenology of Thought?' in *Cognitive Phenomenology*. T. Bayne & M. Montague (eds). (Oxford: Oxford Academic).
- Unger, P. (1979). "Why There Are No People." *Midwest Studies in Philosophy*. Vol. 4
- Unger, P. (1990). *Identity, Consciousness, and Value*. (Oxford: Oxford University Press).
- Vaidya, A. J. (2008). 'Modal Rationalism and Modal Monism,' *Erkenntnis*. Vol. 68. No. 2.
- Van Gulick, R. (1993). 'Understanding the Phenomenal Mind: Are We All Just Armadillos?' In *Consciousness: Philosophical and Psychological Aspects*. M. Davies and G. Humphreys, (eds.). (Oxford: Blackwell).
- van Inwagen, P. (1990). *Material Beings*. (Ithaca and London: Cornell University Press).
- Vintiadis, E. (2019). 'Property Dualism.' In *Introduction to Philosophy: Philosophy of Mind*. H. Salazar (ed). (Rebus Foundation).
- Weatherson, B. (2003). 'Many many problems.' *Philosophical Quarterly*. Vol. 53.

- Weiss, M., Sousa, F., Mrnjavac, N., Neukirchen, S., Roettger, S., Martin, M. (2016). 'The physiology and habitat of the last universal common ancestor.' *Nature Microbiology*. Vol. 1.
- White, S. (1986). 'Curse of the Qualia.' *Synthese*. Vol. 68.
- Whitehead, A. N. (1929a). *Process and Reality*. (New York: Macmillan Publishing).
- Whitehead, A. N. (1929b). *The Function of Reason*. (Princeton: Princeton University Press).
- Whitehead, A. N. (1938). *Modes of Thought*. (New York: Macmillan Publishing).
- Williamson, T. (1994). *Vagueness*. (New York: Routledge).
- Wittgenstein, L. (1973). *Philosophical Investigations*. (New Jersey: Pearson).
- Worley, S. (2003). 'Conceivability, Possibility, and Physicalism.' *Analysis*. Vol. 63.
- Worrall, J. (1989). 'Structural Realism: The Best of Both Worlds?' *Dialectica*. Vol. 43.
- Wright, C. (2001). 'On being in a quandary: Relativism, vagueness, logical revisionism'. *Mind*. Vol. 110.
- Wright, S. (1953). "'Gene and Organism.' Presidential Address to the American Society of Naturalists (Sept. 9, 1952)." *American Naturalist*. Vol. 87.
- Wright, S. (1975). 'Panpsychism and Science.' J.E. Cobb, Jr. and D.R. Griffin (eds.), *Mind in Nature*. (Maryland: University Press of America).
- Wright, S. (c. 1985). 'Psychophysical Identism.' Unpublished manuscript. Accessed through Steffes, (2006).