Against the physicalist zeitgeist: evaluating the contemporary case for physicalism and the prospect of property dualism

by

Ben Hibbert

Submitted in partial fulfillment of the requirements for the Degree of Master of Philosophy in Philosophy

June 2018

Keele University

Abstract

This thesis employs Chalmers' (1996) hard problem of consciousness in an attempt to highlight the fundamental issues with the physicalist zeitgeist that has predominated much of contemporary philosophical thought, before championing an alternative metaphysic, predicated upon a form of monistic property dualism, which may hold the potential to solve the mystery of consciousness without falling foul to the issues faced by physicalism. I begin my inquiry with an explication of those physicalist strategies which have attempted to maintain their metaphysic in light of the hard problem, with a particular focus upon Dennett's (1991) eliminativism, the reductive representationalism held by Dretske (1996) and Tye (2000), and the phenomenal concept strategy established by Loar (1990/9) and Balog (2009/12), before attempting to explicate how all such physicalist strategies are forced to warp our conception of what conscious experience actually is in order to maintain their metaphysic, and contending that, due to the intractable nature of the ontological gap underpinning the hard problem, the contemporary anti-physicalists are right to appeal to a form of ontological dualism that posits consciousness as a fundamental constituent of our reality. From here, I explore the anti-physicalist contentions that have seen the most success in the contemporary literature, and contrast the strengths of substance dualism (Swinburne 1986-2013), panpsychism (Strawson 2006) and a naturalized, or monistic, property dualism (Chalmers 1996), before concluding that, ultimately, both substance dualism and panpsychism contravene upon our understanding of natural laws in such a way that either fractures our evolutionarily constituted worldview (substance dualism), or contradicts our understanding of thermodynamic theory (panpsychism). Thus, I argue, monistic property dualism reveals itself to be the metaphysical framework with the potential to encompass the most explanatorily robust, and metaphysically coherent, solution to the hard problem.

Contents

Introduction **1**

Chapter 1

The Intractability of the Hard Problem **11**
1.1 Formulating the foundations of the problem: the failings of Reductivism *11*
1.2 Strong and Weak Emergence 22
1.3. *New Mysterianism* 29

Chapter 2

A further problem: epistemic justification without phenomenal experience **35**
2.1 The fundamental problem and three potential physicalist solutions *55*

Chapter 3

Dennett's Illusionism **59**

Chapter 4

Reductive Representationalism **76**

Chapter 5

The Phenomenal Concept Strategy **100**

Chapter 6

Towards a paradigm shift: supplanting monistic physicalism with ontological dualism **124**
6.1 Swinburne's Substance Dualism *127*
6.2 Monistic Property Dualism and Panpsychism: towards a potential solution to the Hard Problem 142

Conclusion **168**

Acknowledgments

I wish to express my earnest gratitude for the patience and guidance provided by my supervisor James Tartaglia, whose wisdom has left a lasting impact upon both this thesis and myself. I would also like to thank my partner, Jodie, and my parents, Alison & Philip, for their steadfast love, support, and, most of all, for having endured me amidst long days entrenched in contemplation of being.

Introduction

Sitting here at this point in space-time, I feel the chair beneath me, I feel the soft indentation of my fingers upon the keyboard, I see the letters cascade upon the screen, and I breathe meaning into these letters as I formulate them into words. All of this occurs simultaneously, as a beautiful symphony orchestrated and unified by this 'I' as a subject of experience. These experiences culminate to form the subjective phenomenal character of 'what it is like' (Nagel 1974, p. 442) *for* me as this subject of experience, and it is the fact that there *is* 'something it is like' (ibid., p. 442) for me to subjectively undergo experiential states that typifies my existence most indubitably, for this is the singular phenomenon that is realized with such visceral immediacy that it seems impossible to doubt. Yet, all attempts to employ the physical abrasiveness of quantitative science as a medium from which to reconcile this seemingly unquantifiable phenomenal character have proven redundant. This is the dilemma posed to the physicalist metaphysic. As, upon establishing the brain as a purely structural or functional physical system, physicalism is left incapable of explaining how and why this culmination of physical matter should give rise to my abounding, inner phenomenal experience. This is what Chalmers (1996) terms the 'hard' problem of consciousness[1], and exemplifies a contemporary formulation of a long established 'fundamental problem' with the physicalist metaphysic.

The foundation of this problem can be traced back as far as ancient Greece, and typifies the divide between idealists (see Plato 380 BC, Berkeley 1710) who

[1] Consciousness in this context is concerned with subjective, phenomenal experience. Throughout I shall be referencing consciousness with this in mind, whilst also employing the concept of 'qualia' to denote the subjective, phenomenal properties of experience.

champion the non-physical 'spirit' or 'subjectively experiential mind' as the fundamental property from which all further properties grow[2], and physicalists (see Dennett 1991) that subscribe to what Husserl (1913) references as our 'natural attitude' (p. 57). This natural attitude foregoes the fundamental nature of the mind, and instead proceeds as if the information made available via our phenomenal experience presents external reality as it is in actuality. Husserl (1913) outlines this view as follows:

'I find the actuality as a factually existent actuality and also accept it as it presents itself to me as factually existing. No doubt about or rejection of data belonging to the natural world alters in any respect the general positing which characterizes the natural attitude. The world is always there as an actuality.' (p. 57)

For Husserl (1936), this natural attitude stood in contrast to the 'phenomenological attitude' adopted by the idealists, who posited that our capacity to 'stand above the world' and reduce everything to mental 'phenomenon' (p. 152) stood as evidence for phenomenological experience being the only indubitable actuality. Thus, whilst idealism seemed to stifle epistemic progression[3] via upholding the claim that only consciousness indubitably exists, Husserl's (1913/36) depiction of our 'natural attitude' describes a framework capable of epistemic advancement, via allowing for 'lifeworldy objects' to contain their own spatially extended 'material realities' (Overgaard 2004, p. 20), and exist in a state of physical actuality that is, in itself, more 'factually existent' than the non-physical mind. Over time, the epistemic and

[2] Although in the case of Berkeley (1710) this extends only so far as transcendental idealism, in which one only doubts one's epistemic capacity to know properties beyond the mental.
[3] Here I am employing the concept of epistemic stagnation to denote a state in which an epistemic framework presupposes its own limits, and, as such, inevitably stagnates.

metaphysical commitments of this natural attitude coalesced to form the metaphysic of monistic physicalism, which now dominates the fields of philosophy and natural science. This metaphysic adopts the contention that those phenomena which are spatially and physically quantifiable are more 'factually existent' than their non-physical counterparts, whilst maintaining that, as we only have evidence for the existence of those properties that contain their own 'material realities', we should appeal to an ontological monism, and conclude that our reality only contains one type of thing: the physical. So that the brain, for example, just *is* a spatially extended physical substance, encompassing a set of underlying materials and functions, but withholding no further non-physical 'spirit' or 'mental property' beyond that of which can be physically quantified (this is Dennett's 1991 stance). As such, this physicalist 'natural attitude' simultaneously avoided the epistemic stagnation inherent within idealism, whilst appealing to our intuitive inclination to describe phenomenon presented to us in our sense data as encompassing actually existent physical identities that are ontologically independent of the mind. So, as the antithesis of idealism, this 'natural attitude' became rooted in a monistic physicalism, which posited physical, quantifiable identities as the only 'factually existent actuality', and allowed us to meaningfully communicate about the world by adopting a purportedly 'commonsense' ontological commitment that facilitated physical quantifications and avoided epistemic stagnation.

Thus, the strength of this ontological commitment has been grounded within its capacity to facilitate epistemic advancement, as upon dethroning the mind from the epistemic pedestal championed by the idealists, and embracing our 'natural attitude', monistic physicalism creates a metaphysic, rooted in the contention that only

physically quantifiable phenomena can be said to exist, that allows for the quantification of physical information, and, in turn, provides grounding for all the natural sciences we know today. Indeed, it is the scientific and evidential consistency of this metaphysic that has strengthened the contemporary move to denounce idealism, in favor of the reductive, monistic physicalism that now predominates the current philosophical, and scientific, zeitgeist.

Physicalism, then, presents a robust narrative, and purportedly provides a metaphysical 'key' from which to explain, via causally closed scientific quantification and reduction, all observable, physical substances to further observable, physical substances. However, whilst the metaphysical framework that underpins the natural sciences has undeniably advanced our understanding of the physical, upon employing this same framework in an attempt to explain phenomenal experience, we find ourselves no better equipped than the ancient Greeks. So, natural science, for the most part, has avoided any attempt to reconcile the experiential mind and, as a result, we are confronted with neat, contemporary formulations of this 'hard' problem that, in truth, has plagued our epistemic framework from the very beginning.

The disparity that now lies at the heart of this problem is that the physical sciences, which act as our most championed means of epistemic enquiry, tend to simultaneously rely upon and disregard conscious experience. As, whilst phenomenal experience seems to be epistemologically antecedent to science by grounding our capacity for experiential observation, science's adherence to the physicalist 'natural attitude' often results in the non-quantifiable, or non-physical being denied outright. So, phenomenal consciousness, as an experiential phenomenon, becomes difficult to

reconcile with a philosophical or scientific method underpinned by a physicalist metaphysic that can only affirm the existence of the quantifiable structures or functions of physical phenomena. By means of recompense, physical science either commits itself to the patent absurdity of denying the existence of phenomenal experience entirely (see chapter 3), or attempts to find a means from which to incorporate the phenomenal identity of subjective experience into a framework that begins with physical phenomenon (see chapters 4, 5). These responses all massage the identity of consciousness, in an attempt to physically explain (or eliminate) experience, so that the identity of phenomenal experience is reduced to a given structure or function, which is, in itself, fundamentally physical. As such, these physicalists effectively attempt to deny the import of the problem of consciousness, for if any one of these iterations can show that physicalism is capable of integrating phenomenology (properly defined as an experiential awareness of 'what it is like') into the physicalist metaphysic, then the coherence and metaphysical completeness of science remains unchallenged.

These attempts to maintain the coherence of physicalism shall act as the primary focus of this thesis. I endeavor to highlight the fundamental inadequacies with the physicalist appeal to a 'bottom up' approach, in which the definition of subjective conscious experience is massaged to suit a physicalist framework, and this 'quasi-consciousness' is somehow posited as an emergent feature of a framework that necessarily begins with physical laws (Dennett 1991 seems particularly susceptible to this). My contention throughout shall be that these attempts to massage our definition of phenomenal consciousness result in 'solutions' that fail to truly address the fundamentals of the problem. Further, I shall contend that any future purported

physicalist solution to the problem of consciousness shall face the same fate, as a scientific framework that begins with physical phenomenon will forever remain incapable of employing observation to achieve an understanding of that which makes observation possible. So, just as we do not expect a measuring tool, for example a ruler, to effectively measure itself; we should not rely upon physical observation to effectively explain itself. Instead, we must begin with the mental property that makes observation and explanation possible, and posit this phenomenon as fundamental. Thus, I contend we must allow for further, ontologically fundamental mental phenomena that exist beyond physical phenomena, so that we may glean a means of explaining the property that is simultaneously so fundamental and intractable to science.

As a result, and in line with one kind of contemporary anti-physicalist stance (see Chalmers 1996, Strawson 2006, Nagel 2012, Swinburne 1986-2013), I fall in favour of bridging the gap between idealism and physicalism, so as to posit a metaphysic that allows for consciousness as a fundamental phenomenon, without discounting physicalism entirely. As such, whilst I shall be maintaining that reductive, physical science and phenomenological consciousness seem to encompass contrary metaphysical commitments, it is my contention (along with the contentions of contemporary anti-physicalists, such as Nagel 2012 and Chalmers 2010) that in order to maintain the coherency of our epistemic and metaphysical frameworks, we must integrate experiential, subjective phenomenology into our understanding of the physical brain. Without this integration, we risk either undermining the scientific methods that have had such evidential success in illuminating the existence of the natural laws underpinning our reality, or foregoing the fundamental nature of the

consciousness that antecedes this scientific understanding, and, in so doing, advancing a worldview that necessarily fails to explain that which is most epistemologically fundamental: experience. The problem the physical sciences face, however, is precisely the 'hard' problem iterated by Chalmers (1996), as such a reconciliation seems impossible whilst employing a physicalist metaphysic so constrained by the claim that nothing exists beyond the observable; leaving the physical sciences wholly unequipped to investigate the subjective phenomenology of experience, and resulting in the experiential consciousness that is epistemologically so fundamental to science being discounted. Thus, in the course of this thesis, I shall uphold the central claim that, in order to maintain a coherent worldview, we must cease to exclude phenomenology from our epistemic framework, and, as such, must investigate metaphysical frameworks beyond that of physicalism, that may allow for the reconciliation of the physical brain with the subjective character of experience.

In order to uphold an appeal to shift the contemporary zeitgeist, I must first address the physicalist theses that deny the import of the problem of consciousness; for, in order for my thesis to stand, it must be shown that current physicalist attempts to reconcile consciousness are flawed. In chapter 1, I ground the 'hard' problem, and employ the contentions of 'new mysterianism' (see McGinn 1989) as a means to explicate the deeper issues underpinning any attempt to employ physical science as a means from which to explain consciousness. In chapter 2, I employ these contentions to unravel an epistemic vacuity that, potentially, permeates the entirety of the physicalist framework. From here, I move on to challenge the most robust of the physicalist attempts to maintain their metaphysic via either eliminating phenomenal consciousness, positing it as reducible to physical phenomena, or one that supervenes

on such phenomena. Here I focus upon eliminativism, representationalism and the phenomenal concept strategy (see chapter 3, 4, 5), in the hope that, upon delineating the deficiencies within these frameworks, I may show that phenomenal consciousness cannot be eliminated, posited as an intentional, physical interaction between the brain and its environment, or reduced to a conceptual 'trick' that occurs upon the brain referencing itself. In chapter 3, I highlight the self-defeating nature of Dennett's (1991) attempts to discount phenomenal experience entirely. Within chapter 4, I focus upon Tye's (2000) PANIC thesis as a means from which to typify the broader representationalist contentions, before advancing an argument for the incompatibility of phenomenal character with the externalism that underpins representational accounts such as Tye's. Whilst chapter 5 formulates the current debates and motives underlying an appeal to the 'phenomenal concept strategy', before falling firmly in line with the contention that a formation of this kind is doomed to fall back on ontological dualism. By the close of chapter 5, I hope to have outlined the current motives and articulations underlying the metaphysical doctrine of physicalism, whilst also advancing arguments capable of explicating the fundamental flaws inherent within current attempts to employ physicalism as a means from which to solve the 'hard' problem.

With chapters 3, 4 and 5 as grounding, I move on to assert that if physicalism demands that we disregard conscious experience entirely (eliminativism), posit phenomenal content as one and the same as the properties of external, physical objects (representationalism), or posit that any notion of an ontological gap is essentially just a conceptual misunderstanding that occurs upon the brain referencing itself (phenomenal concept strategy), then the cogency of physicalism lies precariously

upon a series of empirically unfounded, highly counter-intuitive propositions. With the problems inherent to the physicalist metaphysic highlighted, within chapter 6 I investigate the contentions of contemporary anti-physicalists, who have attempted to advance a paradigm shift away from the physicalist zeitgeist that has so predominated modern metaphysics. In the hope that, by investigating these competing metaphysical doctrines, we may unravel a framework with the potential to resolve the problem of consciousness in a manner that both coherently captures phenomenal experience in entirety and remains coherent with our otherwise robust worldview. This resolution, the anti-physicalists contend, is found upon denouncing the monism of physicalism, and positing mental phenomenon as an irreducible property or substance that exists alongside physical properties as a fundamental constituent of the universe. With particular focus upon Swinburne's (1986-2013) substance dualism, Strawson's (2006) panpsychism, and the monistic property dualism endorsed by Chalmers (1996), I articulate the current debates underpinning these metaphysical doctrines, before presenting a novel case for the inability of both substance dualism and panpsychism to be reconciled with the natural laws underpinning our known reality, and, ultimately, championing a form of monistic property dualism, of the kind upheld by Chalmers (1996), which is capable of providing a solution to the hard problem whilst remaining entirely consistent with said laws.

Thus, I conclude that upon positing a neutral monism that leads to property dualism, we reveal a metaphysical framework with the potential to unify the phenomenal character of experience with the physical brain, and, in so doing, provide a means from which to save philosophical enquiry from the intractability of consciousness. Thus, I posit that a potential 'solution' to the problem of consciousness is found upon

establishing mental properties as a fundamental kind of property, and contend that, without a metaphysical solution of this ilk, experience shall forever remain beyond reconciliation with our physical understanding. In order to reach this conclusion, however, it seems judicious to first address the underlying propositions that have resulted in the contemporary preoccupation with this ‘hard’ problem.

Chapter 1

The Intractability of the Hard Problem

1.1 Formulating the foundations of the problem: the failings of reductivism

Why should a physical system, no matter how complex and well organized, give rise to experience at all? Why is it that all this processing does not go on 'in the dark', without any subjective quality? Right now, nobody has good answers to these questions. This is the phenomenon that makes consciousness a real mystery.

Chalmers (1996, p. 359)

Upon posing this question, Chalmers (1996) effectively reignited the age-old philosophical inquiry into the failings of physicalist metaphysics. As, with deft simplicity, Chalmers highlights that the same physicalist metaphysic we employ to achieve scientific success is incapable of explaining the phenomenal feel of 'what it is like' that appeals to our most visceral intuition. Thus, in formulating his 'hard' problem, Chalmers rekindled the divide between idealism and the scientific natural attitude, via charging physicalism with an explanation for our seemingly indubitable phenomenal experiences.

The intricacies of Chalmers' formulation take foundation in the failings of physical reductivism, which asserts that phenomenon 'Q' can be explained, via causal reduction, by the more fundamental phenomenon 'P', 'so that P = Q' (Blamauer 2011, p. 100). The issue, Chalmers (1996) purports, is the incapacity to identify conscious experience in the same manner used to identify physically reductive properties. Such that, whilst the reduction of water (Q) to the physical properties of H2O (P) is possible, a similar reduction cannot be employed to account for conscious experience

(Q) within the physical properties of a neurologically constituted brain-state (P). Chalmers (2003) contends this issue arises because reductive explanations of physical phenomenon are only capable of accounting for the 'structure and function' (Chalmers 2003, p. 103) of physical properties, without ever addressing the phenomenal feel of 'what it is like' to have an experience. Thus, the scientific 'natural attitude', in its adherence to quantifiable objectification, fails to physically explain the experiential identity that lies at the heart of what it means to *be* this human subject.

Chalmers (2003) appeals to a conceivability argument in order to compound his point, in which he highlights that, if we uphold the physicalist notion that experience *can* be accounted for within the identity of physical structures or functions, then a physically identical subject to me must necessarily share my phenomenological state. As Kripke (1971) attests: 'for any objects *x* and *y*, if *x* is *y*, then it is necessary that *x* is *y*' (p. 137). So, if physicalism attempts to reduce phenomenal experience to a physical phenomenon, it must hold that phenomenal experience necessarily shares the identity of the lower-order physical phenomenon to which it is causally reducible. However, as Chalmers (2003) highlights, it is not difficult to conceive of my physical clone *not* having the phenomenal 'feel' of what it is like, but instead existing as nothing beyond a functioning, physically structured substance occupying a certain spatiotemporal point (p. 105). Hence, whilst we could not conceive of water being anything other than H20 without changing our definition of 'water' (Putnam 1973, p. 701), we *can* conceive of a 'philosophical zombie' (Chalmers 2003, p. 105) withholding our exact physiological makeup, but being devoid of the phenomenon we reference as phenomenal consciousness. The problem becomes that, if this is at least conceivable, there can be no necessity in the relationship between the identities of

phenomenological consciousness and the physical brain, as we can conceive of one occurring without the other. So, whilst the conceptual identity of water employed on planet earth can never not be H20, we *can* hold the concept of a brain not withholding consciousness, and, as such, we *can* posit a distinction between the identities of phenomenal consciousness and the physical brain. Thus, Chalmers (2003) concludes that if we were to posit 'P' as the 'conjunction of all microphysical truths about the universe' (p.106), in which every 'fundamental feature of every fundamental microphysical entity' (p. 106) was specified, we would still be left with the problem of explaining the truth that we hold phenomenal consciousness (Q), because if we can conceive of a non-phenomenal 'philosophical zombie', then *P&~Q* is conceivable, and, therefore, it is metaphysically possible for Q to exist as an ontological identity that is irreducible to P. As such, if this is conceivable and therefore (metaphysically) possible, the physicalist notion that, by a matter of metaphysical necessity, all phenomena is reducible to a more fundamental microphysical truth, is rendered false by virtue of a phenomenal truth (Q) that presents an ontological identity beyond the scope of microphysical truths.

The physicalist retorts to this line of argument are abundant (see Dennett 1998, Yablo 1993, Ashwell 2003), with all presenting categorical rejections of Chalmers' inference from conceivability to (metaphysical) possibility, and, in turn, presenting a case for an outright rejection of Chalmers' (1996/2003) original, more simplistic, formulation of his conceivability argument, which is articulated as follows:

(1) *P&~Q* is conceivable

(2) If *P&~Q* is conceivable, *P&~Q* is metaphysically possible

(3) If *P&~Q* is metaphysically possible, materialism is false.

(4) Materialism is false.

These responses hinge on the notion that Chalmers' formulation presumes the validity of either the epistemic thesis of what can be conceived as expressed in premise 1 (Dennett 1998, Ashwell 2003), or the leap from epistemic conceivability to modal possibility as evidenced in premise 2 (Yablo 1993). Whilst the former attempts to deny our capacity to even conceive of (or imagine) philosophical zombies by maintaining that those philosophers who claim zombies are conceivable 'invariably underestimate the task of conception (or imagination)' (Dennett 1998, p. 172), the latter endeavors to mount a direct attack upon the Humean (1968) notion 'that nothing we imagine is absolutely (metaphysically) impossible' (p. 32). With this in mind, Chalmers (2010) attempts to explain away the issue highlighted by Dennett (1998), and address the stronger argument that denies the move from conceivability to metaphysical possibility. In order to achieve this, Chalmers (2010) extrapolates a more robust conceivability argument by way of delineating between the prima facie conceivability of a hypothesis that is plausible by mere initial consideration or a priori reasoning, but is directly 'tied to a subject's contingent cognitive limitations' (p. 143), and the 'ideal conceivability' of a hypothesis that cannot be denied even upon employing an 'ideal rational reflection' (p. 144), which 'abstracts away from those (cognitive) limitations' (p. 229). Chalmers (2010) attempts to clarify this distinction by employing the following examples:

'(1) '2+2=5' is neither prima facie conceivable nor ideally conceivable;

(2) Where *S* is a highly complex but provable mathematical truth, ~*S* will be prima facie conceivable for most subjects, but it is not ideally conceivable;

(3) Where *S* is 'There is a flying pig', *S* is prima facie conceivable, and is almost certainly ideally conceivable.' (p. 143)

With these examples in mind, we can simplify this distinction as follows: prima facie conceivability occurs upon a subject imagining a case in which the initial hypothesis for S's existence is taken to be coherent, whilst ideal conceivability occurs when this initial prima facie conceivability cannot be logically shown to contradict upon employing 'ideal rational reflection' (Chalmers 2010, p. 144). Yablo (2008) provides some explanatory weight to this distinction by classifying, what Chalmers (2010) might term, prima facie conceivability as indicative of a 'believability sense' (p. 50), and 'ideal conceivability' as a sense in which we are able to present 'the appearance of possibility' (p. 50). Such that, in the case of Chalmers' highly complex but provable mathematical theorem S, we may be able to hypothesize using prima facie conceivability that we believe ~*S* to be the case, but upon employing ideal rational reflection, we reduce the possibility of ~*S* being true to such an extent that we render our prima facie belief redundant, and employ ideal conceivability to affirm S's existence indubitably. In this example, what Chalmers terms a 'ideal negative conceivability' (p. 144) is employed, in which the conception of S is established and deductive logic leads us to the impossibility of ~*S*. However, Chalmers (2010) also posits its opposite, what he terms an 'ideal positive conceivability' (p. 144), in which the conception of, for example, philosophical zombies (Z) is established, and

deductive logic leads us to affirm the (metaphysical) possibility of Z, because the prima facie belief can *not* be 'defeated upon ideal rational reflection' (p. 144)[4].

Chalmers (2010) argues that critics (see Dennett 1998, Ashwell 2003), who contend that the ideal positive conception of zombies is inconceivable on the grounds that imagining an amalgamation of the brain's physical processes also involves imagining conscious processes, are flawed by virtue of their over-reliance upon the prima facie belief that conscious and physical processes are necessarily intertwined. Indeed, as Chalmers (2010) argues, whilst a prima facie belief about the relation between phenomenal and physical processes may lead the physicalist to deny the notion of prima facie zombie conceivability, it remains perfectly possible for the physicalist to employ ideal rational reflection in order to ideally and positively conceive of cases in which the absence of this relation is a metaphysical possibility, and so why should the physicalist face a difficulty in imagining the metaphysical possibility of a case in which the extrinsic, functional or spatial properties of physical processes are absent of intrinsic phenomenal properties? And, as we *can* conceive of this (regardless of our prima facie beliefs)[5], the question now becomes whether or not this ideal conception is itself enough to ground metaphysical possibility.

Chalmers (2010) argues that this is the case, and highlights that instances in which the leap from conceivability to (metaphysical) possibility is denied (See Yablo 1993)

[4] 'Where *S* is 'There is a flying pig', *S* is prima facie rational conceivable, and is almost certainly ideally conceivable.' (p. 143) would be an example of this, as we cannot employ ideal reflection to indubitably affirm that the hypothesis is (metaphysically) impossible (in all possible worlds).

[5] Worley (2003) denies this claim by contending that ideal rational reflection would imply a greater breadth of scientific and logical reasoning than that of which we are currently privy to, and so concludes that an ideal conception of zombies is impossible. However, Chalmers (2010) refutes this by highlighting that, in order for zombies to be beyond the scope of ideal reflection, phenomenal concepts must be functional concepts, and, as shall be explored in chapter 5, this seems difficult to establish.

correctly focus upon the difficulty of leaping from prima facie conceivability to (metaphysical) possibility, whilst failing to account for the inference from ideal positive conceivability to metaphysical possibility. Chalmers frames this argument by establishing a distinction between primary and secondary conceivability, in the hope of avoiding Putnam's (1975) dismissal of a posteriori conceivability arguments on the grounds that just because 'we can perfectly well imagine having experiences that would convince us (and that would make it rational to believe) that water is not H20' (p. 233), it does not mean that this is metaphysically possible, as 'once we have discovered that water (in the actual world) is H20, nothing counts as a possible world in which water isn't H20' (p. 233) . The argument follows that whilst secondary conceivability is achieved from a posteriori, empirical factors, which entrench us in prima facie beliefs and leave us open to Putnam's objection; primary conceivability is achieved by way of pure a priori, ideal reasoning, and, as a result, avoids Putnam's dismissal of a posteriori conceivability. Such that, whilst Chalmers concedes that the physicalist can deny the possibility of zombies 'on the grounds of an a posteriori identity between phenomenal and physical properties' (Chalmers 2010, p. 308), he maintains that we must detach ourselves from the limitations of our prima facie beliefs gleaned by way of empirical factors, and, as such, urges that we must place impetus unto the possibility of ideal, a priori, primary conceivability leading to metaphysical possibility.

In order to establish how this may be so, Chalmers (2010) maintains that upon employing ideal, primary conceivability to, for example, imagine that water is not H20, we are in fact distancing ourselves from the prima facie belief that 'water is H20' by employing the primary concept of 'water' to denote any referent with 'water-

like' qualities, and it is this primary conceivability that allows us to imagine the metaphysical possibility of a Twin-Earth in which oceans and lakes are comprised of XYZ. Upon employing this ideal, primary conception of 'water', Chalmers (2010) argues that 'our conceiving involves access to a possible world' (p. 146), in which there is 'a link between primary conceivability and metaphysical possibility' (p. 146), because, whilst 'water is not H2O' is not secondarily conceivable, or a posteriori true, for us as centered beings at this point in space-time on Earth, it *is* secondarily conceivable, and a posteriori true, for our opposite centered at this point in space-time on a possible Twin-Earth. This is because, upon employing the primary intension of 'water', we are a priori referencing stuff with water-like qualities in all possible worlds; such that the truth of the concept is not limited to the world, but is instead limited only by the beliefs of the individual centered in said world. Meaning that whilst the employment of the a posteriori, secondary intension of water as H20 leads us to reference water on Earth, we are still able to detach ourselves from this a posteriori, prima facie understanding of 'water' in order to infer that, if our primary intension of 'water' is simply picking out any referent with watery qualities, then 'water' picks this out regardless of where we are centered. As Kallestrup (2012) points out: 'only when the microstructure of water was discovered did 'water' become a natural kind term expressing a natural kind concept. The mistake is to impose scientific intuitions about concept individuation on those who possess concepts pre-scientifically', such that If H20 is simply the 'watery stuff' on earth, then 'water' picks out H20, if XYZ is the 'watery stuff' on Twin-Earth, then 'water' picks out XYZ (p. 67). Similarly, whilst there may be an a posteriori link between physical and phenomenal properties on earth, if the primary intention of 'phenomenal property' is simply picking out a referent with phenomenal qualities, then it is conceivable that

there exists a possible world in which there is no such link. With this established, Chalmers (2010) argues:

'We can say that when the primary intension of S is true at some centred world (i.e., when some centred world verifies S), S is primarily possible, or 1-possible. When the secondary intension of S is true at some world (i.e., when some world satisfies S), S is secondarily possible, or 2-possible. Then 'water is not H_2O' is not 2-possible, but it is 1-possible.' (p. 147)

With this as his grounding, Chalmers (2010) now offers a reformulation of his conceivability argument, in the hope of avoiding the more simplistic physicalist arguments against the inference from conceivability to (metaphysical) possibility:

'(1) *P&~Q* is conceivable

(2) If *P&~Q* is conceivable, then *P&~Q* is 1-possible

(3) If *P&~Q* is 1-possible, then *P&~Q* is 2-possible or Russellian monism is true.

(4) If *P&~Q* is 2-possible, materialism is false.

(5) Materialism is false or Russellian monism[6] is true.' (p. 152)

Thus, as we can conceive of the conjunction of microphysical truths (P) without phenomenal experience (Q), it is primarily conceivable that *P&~Q* is primarily possible in some possible world. If it is primarily possible, it is therefore secondarily

[6] Russel's (1927) Russelian Monism can be construed as a form of monistic physicalism that is closely related to both property dualism as well as panprotopsychism, and simply affirms: (1) Perception and science exclusively denote the structure or function of physical entities without picking our their intrinstic nature, and (2) An explanation of phenomenal consciousness will only be achieved upon understanding the intrinsic nature of physical entities, and so, as we can not achieve this by empirical means, we can, at best, only speculate about the intrinsic nature of microphysical properties.

possible that some possible world empirically satisfies the existence of philosophical zombies, and, therefore, physicalism is false. Or, the a posteriori, empirical, scientific approaches that have led us to embrace physicalism on this possible world are simply ineffective means of delineating the 'special nature' of the microphysical properties that are 'tied to consciousness' (Chalmers 2010, p. 152) in all possible worlds, and, therefore, Russelian Monism[7] is true.

The consequence of this argument is, therefore, twofold: either physicalism is rendered false by virtue of our capacity to primarily conceive of philosophical zombies, or physicalism is shown to over-exert its explanatory power by virtue of inviting the modal fallacy that occurs upon relying on the empirical facts in this possible world to infer the intrinsic nature of microphysical properties in all possible worlds. Thus, according to Chalmers, physicalism, and, in turn, the scientific 'natural attitude', in its adherence to quantifiable objectification, fails to account for, or presumes the non-existence of, the experiential identity that lies at the heart of what it means to *be* this human subject.

As such, I contend that the really important question that Chalmers' appeal to philosophical zombies highlights is: how can physicalism, whilst employing the scientific method, delineate the intrinsic nature of microphysical properties in all possible worlds in such a way that explains why philosophical zombies are impossible? As, it now seems clear that upon defining physical properties as

[7] I note here, in line with Chalmers (2010), that whilst Russelian Monism may be construed as a form of physicalism, 'it relies on speculation about the special nature of the fundamental properties in microphysics', and so 'has much in common with the property dualism that physicalists will want to reject' (p.152).

encompassing a set identity, which exclusively explains their quantifiable function, structure or position in space-time, the onus of proof lies with the physicalist to explain how these purely physical properties may account for the experiential identity of consciousness, so as to elucidate, in a manner consistent with physicalism, an explanation for both *how* a physicalist theory of consciousness may account for our seeming to contain non-physical, experiential properties that account for our being experientially conscious, and *why* such a theory necessarily precludes the possibility of philosophical zombies. I argue an explanation of this kind must be found within the physicalist theories themselves, and so, in the subsequent chapters (see 3, 4, 5), I look to these frameworks to establish a theory of consciousness capable of physically accounting for experiential identities.

The inevitable problem faced by these theories is one of reconciling experiential identities with the physical observations employed by physicalism. This is because, as Jackson's (2003) infamous knowledge argument highlights, experiential information remains epistemologically distinct from (and inexplicable with) physical information, as any *a priori* accumulation of physical information pertaining to the functions or structures of 'redness', for example, will forever remain incapable of accounting for the new *a posteriori* knowledge gained upon learning 'what it is like' to phenomenally experience 'redness' for the first time. As such, it seems physicalism, in its attempts to 'observe' consciousness by appealing to physically quantifiable neurological structures and functions of the brain, must first provide an account capable of fully explaining the experiential, phenomenal information that underlies our capacity for observation. Thus, as physical reductivism sticks strictly to the schema 'all entities are nothing over and above physical entities' (Wilson 2005, p.

426), the notion of conscious experience charges physicalism with a metaphysical and epistemic quandary, which seems to appear insolvable without either: denying that this phenomenon contains the experiential identity we have ascribed to it (see chapters 3, 4, 5), denying the monism of physicalism by positing the existence of an experiential property that is ontologically irreducible to physical substances or properties, (see chapter 6), or finding a means from which to explain this property that *is* seemingly 'over and above physical entities' without foregoing the fundamental nature of physical substances.

1.2 Strong and Weak Emergence

Emergentism arose as one such attempt to explain experience without foregoing the physicalist paradigm completely. This framework aims to explain how novel properties, such as consciousness, may emerge from underlying, purely physical, correlates. The concept of emergence can be articulated in two distinct forms: 'strong emergence' of the kind upheld by Alexander (1920), and 'weak emergence' of the kind upheld by Searle (1992). In what follows, I explicate what is meant by these concepts, before addressing the distinction between these two forms of emergentism, and, ultimately, concluding that both forms fail to explain the emergence of phenomenal experience.

Chalmers (2008) defines the distinction between these two concepts as follows:

Strong Emergence: *'A high-level phenomenon is strongly emergent with respect to a low-level domain when the high-level phenomenon arises from the low-level domain,*

but truths concerning that phenomenon are not deducible even in principle from truths in the low-level domain.' (p. 244)

Weak Emergence: *'A high-level phenomenon is weakly emergent with respect to a low-level domain when the high-level phenomenon arises from the low-level domain, but truths concerning that phenomenon are unexpected given the principles governing the low-level domain.'* (p. 244)

As such, both notions of emergence are predicated upon the notion that an emergent property X is individuated as having a property of something new, such that this property confers a 'new identity on the thing that has it' (Wyss 2004, p. 3). However, whilst strong emergence entails that the novel phenomenon 'X' may emerge from the low-level phenomenon 'Y' even if the type or token identity of X is in no way 'downwardly' causally deducible from the identity Y, weak emergence entails that the novel phenomenon X may emerge from low-level phenomenon Y only if the identity of X can be deduced from the underlying identity of Y. Thus, simply, whilst, in the case of strong emergence, we could *not* deduce the emergence of X from Y, we *could*, in the case of weak emergence, deduce that X may emerge from Y.

This notion of weak emergence was championed by the likes of Searle (1992), who likened the novel property of 'liquidity', which emerges upon combining 2 hydrogen atoms with 1 oxygen atom so as to form water molecules, and then loosely bonding these molecules, at a certain temperature, so that they may 'slide past one another' (Strawson 2008, p. 18), to the phenomenal feel of experience that emerges from the underlying neuronal structure of the brain. Thus, just as the individual atoms that

make up water do not account for its liquidity, Searle argues that consciousness cannot be accounted for by individual neurons. Instead, consciousness can only be accounted for by the structuring of said neurons, so that consciousness becomes a 'higher-level or emergent property of the brain', just as 'liquidity is a higher-level emergent property of H2O molecules when they are, roughly speaking, rolling around on each other (water)' (Searle 1992, p. 14). As such, Searle's formation posits consciousness as a 'causally emergent system feature' of micro-level substances (neurons) interacting in such a manner so as to produce a macro-level physical substance (the brain) capable of accounting for the 'higher-level emergent property' of consciousness (Searle 1992, p. 111). However, Searle is careful to explain that whilst the causal interaction between consciousness and the macro-level brain is explicable; we would need 'some additional account of the causal relations' at a micro-level in order calculate consciousness using the physical structures of neurons (Searle 1992, p. 112).

Herein lies the issue with both weak emergentism and Searle's formation, as whilst the liquidity of water can be deduced from the identity of the atoms underlying H20, such that we may logically predict the emergence of liquidity from these atoms, it remains clear that 'neurons are incapable of predicting the property of consciousness' (Havlik 2012, p. 41). As such, Searle seems to be relying upon the extrinsic, interactional, or, in this case, behavioural properties of neurons to account for the causal relations between them, and, in turn, account for the emergence of consciousness. However, it seems very difficult to maintain an appeal to weak emergence whilst upholding the claim that the physical behaviour of neurons produces the experiential sensation of 'what it is like' to undergo such behaviour, as

the behaviour itself remains bound to the physical properties of neurons. Indeed, it seems impossible to explain why the behaviour of a certain structure of neurons, containing a strictly non-experiential physical identity, should give rise to the subjective experiential quality of 'what it is like', because, unlike the physically quantifiable and deducible interaction between H20 and liquidity, the experiential identity of consciousness is not reducible to or deducible from either the physical identity, or the physical structures and/or behaviours of neurons (this particular problem typifies Levine's 1983 'Explanatory Gap'). In turn, Searle's argument for weak emergence seems flawed. As, whilst we *can* employ a weak downward causation to deduce that the novel emergence of the physical, macro-level phenomenon of liquidity is likely to arise from the physical, micro-level atoms of H20, which do not, in themselves, contain the identity of liquidity, we *can not* employ the same weak downward causation to deduce the novel emergence of a non-physical macro-level phenomenon, such as phenomenal experience, from a purely physically constituted, neurological micro-level substance. Strawson (2008) exemplifies this point as follows:

'We can easily make intuitive sense of the idea that certain sorts of molecules are so constituted that they don't bind together in a tight lattice but slide past or off each other (in accordance with van de Waals molecular interaction laws) in a way that gives rise to—is—the phenomenon of liquidity…We [start] in a small set of conceptually homogeneous shape-size-mass-charge-number-position-motion-involving physics notions with no sense of puzzlement…. [And] using the notion of reduction in a familiar loose way, we can say that the phenomena of liquidity reduce without remainder to shape-size-mass-charge-etc.' (p. 61)

Hence, unlike phenomenal experience, liquidity is still a fundamentally physical phenomenon, which is *non-mysteriously* explained by virtue of micro-level physical properties, of a particular physical shape, size or mass, 'sliding past' one another so as to create the physical phenomenon we reference as 'liquidity'. This is quite distinct from phenomenal experience, as the physical identity of neurons (X) is such that we cannot, in any way, deduce the novel emergence of a non-physical experiential identity (Y) arising from the low-level properties of X. Thus, as Benovsky (2015) formulates: the crucial distinction is that 'liquidity of water is explained by the features H2O molecules do have, [so that] it is wholly dependent on these non-liquid features, in a *non-mysterious* way' (p. 344), however, the experiential consciousness of the brain is not explained by the underlying physical properties of neurons, and so consciousness seems to be dependent upon these non-phenomenal properties in a way that is entirely *mysterious.* As such, the crucial point here is that in order for us to deduce non-physical, phenomenal experience Y emerging from physical neurons X, we must infer that the physical identity of X is entirely non-mysteriously essential to the non-physical identity of Y in the same way that the physical identity of H20 is entirely non-mysteriously essential to the physical identity of liquidity. However, it remains clear that this is not the case in regards to conscious experience, and, as a result, it remains clear that we cannot rely upon a form of weak downward causation to account for the emergence of experience. As Strawson (2008) attests:

'For Y truly to [weakly] emerge from X is for Y to arise from or out of X or be given in or with Y given how X is. Y must arise out of or be given in X in some essentially

non-arbitrary and indeed wholly non-arbitrary way. X has to have something – indeed everything – to do with it.' (p. 66)

Under this strict definition of weak emergence, we witness that, unlike in the case of liquidity, in which the physical identity of the atoms of H2O have *everything* do with the physical identity of this emergent phenomena; in the case of phenomenal consciousness (Y), it becomes very difficult to explain how the underlying, entirely non-phenomenal, physical identity of neurons (X) had anything at all to do with the entirely non-physical, phenomenal identity of the emergent property Y. Thus, whilst, in the case of liquidity, we are still discussing a physical identity, such that the physical identity of the emergent property is deducible from the underlying physical identity of atoms; in the case of phenomenal experience, the emergent, experiential identity is novel, and, apparently - unless a physical theory can show otherwise - entirely non-physical, and therefore, non-deducible from the underlying physical identity of neurons. Indeed, 'given how X is', it seems entirely logical to infer that we may *never* have deduced Y from X. At this stage, it seems the weak emergentist must either discount the purely physical identity of X so as to accommodate for the emergence of Y (and in turn discount physicalism via positing a fundamental experiential property capable of accounting for the novel emergence of Y, see chapter 6 for an explication of this strategy), or fall back onto a notion of strong emergence, which is capable of accounting for novel properties, such as Y, emerging from underlying low-order phenomenon X, in a manner that does not require the identity of X to in any way explain the existence of a novel emergent identity Y.

Thus, the physical emergentist must appeal to a more radical form of 'strong' emergentism, in which notions of 'weak downward causation[9]' (Chalmers 2008, p. 250) are modified, and consciousness is said to emerge as a novel property, which is neither deducible from, nor ontologically reducible to, the physical properties from which it ostensibly emerged. This is the stance established by the British emergentists (see Mill 1872, Morgan 1923, Alexander 1920 and Broad 1925), and suggests that emergent properties are not deducible from their underlying physical system. This stance clearly avoids the issues of weak emergentism, as it does not require the underlying physical system to account for the formation of novel identities, and instead employs a 'strong downward causation', in which 'the causal impact of a high-level phenomenon on low-level processes is not deducible even in principle from initial conditions and low-level laws' (Chalmers 2008, p. 250). Thus, low-level properties do not causally or conceptually necessitate high-level properties and, as such, are not required to explain novel emergence. This, I contend, simply represents a clumsy means of escaping the problem faced by Searle, as the strong emergentist is demanding that we not only accept the possibility of non-experiential physical properties spontaneously producing experiential properties, but that we also accept that the underlying lower-order physical properties should not be held accountable for an explanation of how this is possible. Thus, I agree with Strawson (2008) that strong emergentism of this kind seems to demand that we accept the 'brute', impossible occurrence of something spontaneously emerging from nothing, and in so doing, seems to rely upon a 'magic passage across the experiential/non-experiential divide' (Strawson 2008, p. 70).

[9] A 'weak' downward causation of this kind is the sort that weak emergentism employs, and is defined as: 'a causal impact of the high-level phenomenon is deducible in principle, but is nevertheless unexpected' (Chalmers 2008, p. 249).

As a result, we are left incapable of reconciling the place of consciousness within a conceptual framework so entrenched in the physicalist metaphysic. As it seems, in order to account for experience, the physicalist must either warp the experiential identity of consciousness to meet a framework that can only account for structural, functional identity, or rely upon a miracle to explain how non-experiential properties spontaneously produce experiential properties capable of accounting for the phenomenal 'what it is like' to *be* a physical system. Thus, the real problem is that experience represents something that we know exists, but cannot place within our current explanation of the world (Tartaglia 2015, p. 84), and so we are left either attempting to adapt our metaphysical framework to accommodate (see Chapter 6) for experiential properties, or relying upon the same 'natural attitude' that arguably could be blamed for the problem in the first place (see chapters 3, 4, 5). Although, as shall be exemplified in the subsequent section, in this latter case, the problem of attempting to employ observation to explain itself seems truly intractable.

1.3. New Mysterianism

'There is no guarantee that our cognitive powers permit the solution of every problem we can recognize.'

(McGinn 1989, p. 353)

Throughout history this 'hard' problem has been iterated in a variety of formulations (for examples see Leibniz 1714, Locke 1690), but, simply, as articulated in the previous section, it is the problem of explaining how a distinct experiential identity

‘Q’, replete with the quality of phenomenal ‘feelings’, may arise as a result of a physical, non-experiential, ‘feelingless’ identity ‘P’. Whilst this articulation certainly exemplifies the ‘hardness’ of the ‘hard’ problem, ‘New Mysterians’, such as McGinn (1989/95), attest that the intractability that lies at the heart of this problem derives from our inability to explain how ‘something essentially non-spatial emerged from something purely spatial’ (McGinn 1995, p. 101). Arguably, explaining this divide is the fundamental difficulty faced by those confronting the hard problem, and has caused this sub-group of philosophers to deny the possibility of the human species ever having the cognitive capacity to explain experience (these are the contentions that underpin the ‘New Mysterian’ movement see McGinn 1989/1995, Fodor 2001, James 1896).

The problem, according to these philosophers, is that our preoccupation with a science that only explains spatial phenomena leaves any notion of non-spatial phenomena looking ‘more like magic than a predictable unfolding of the natural law’ (McGinn 1995, p. 101). Thus, McGinn (1995) argues for the need to formulate a ‘new conception of space that can overcome the impossibility of finding a place for consciousness in it’ (p. 107), but contends that a conception of this kind will forever elude us as it demands a ‘adequate articulation of consciousness’ (p. 160) that is beyond the scope of human knowledge. As such, McGinn (1995) contends that whilst space *can* accommodate for consciousness, the human mind is not capable of the same accommodation. Thus, the hard problem is not a problem that exists beyond the confines of human reasoning; it is, instead, a natural consequence of a species devoid of the ‘epistemic tools’ (Northoff 2014, p. 220) required to grasp the fundamental nature of reality. This, according to McGinn (1995), occurs because our

understanding of experiential phenomena is limited by our capacity for observation, meaning that any attempt to employ observation to explain itself will fail to confront the 'inner constitution' (p. 108) of consciousness. In this sentiment, I am in agreement with McGinn, as it seems clear that just as a measuring implement fails in effectively measuring itself, I will forever remain incapable of employing observation to explain the phenomenal, subjective 'feel' of 'what it is like', as 'any observation that I might care to make is itself that which was supposed to be observed' (Searle 1992, p. 99).

Whilst such sentiments serve in typifying the difficulties faced by a physicalist solution to the 'hard' problem, as it seems clear that a metaphysical framework predicated on spatial observations shall forever remain incapable of advancing our understanding of the experiential, non-spatial property of consciousness. My point of departure with McGinn (1989) rests in his jump from the intractability of observation employing observation to explain itself, to his formulation of our purported complete 'cognitive closure' to this particular problem, which states:

'A type of mind M is cognitively closed with respect to a property P (or a theory T) if and only if the concept-forming procedures at M's disposal cannot extend to a grasp of P (or an understanding of T).' (p. 350)

Here, McGinn is not simply contending that a physicalist metaphysic reliant upon spatial, physical observations will never explain our capacity for experiential observation (as is my thesis), but instead that a complete understanding of the causal origins of consciousness is cognitively closed to us. The core of this cognitive closure, according to McGinn (1995), rests upon the distinction between our

introspective, non-spatial concept-forming capacities, which reveal consciousness 'from the inside' (p. 100), and our spatial, perceptual concept-forming qualities, which attempt to observe consciousness from the outside. Thus, McGinn does not uphold the dubious contention that introspective consciousness is unknowable to the mind, but instead is contending that an understanding of the causal origins underlying conscious property 'P' would require a reconciliation of these seemingly non-spatial introspective concept-forming capacities alongside our spatial, perceptual concept-forming strategies, so as to unite introspective concepts, which internally capture consciousness, with perceptual concepts, which attempt to externally locate consciousness. The problem with this, according to McGinn, is that the understanding conferred by introspective concepts cannot be brought in line with the understanding conferred by perceptual concepts, as we are devoid of the cognitive tools required to simultaneously reconcile introspective consciousness 'from the inside' whilst employing perceptual concepts to capture consciousness 'from the outside'. Indeed, whilst we can form introspective concepts about consciousness, we cannot employ spatial concepts to simultaneously 'articulate the natural constitution of what we are thinking about' (McGinn 1995, p. 107) and so, in this respect, perceptual and introspective concepts seem entirely cognitively isolated. As such, McGinn concludes that the mind is incapable of shedding light unto the causal origins of non-spatial property (P), because, in order to achieve this, we would need concept-forming capacities that are simultaneously introspective and perceptual, and as a matter of natural fact, we lack such capacities. Hence, McGinn (1989/1995) foregoes any notion that the problem is 'hard' but potentially solvable, and instead maintains that the problem *cannot* be solved due to our epistemic incapacity to even grasp non-spatial concepts in a manner that connects them with spatial concepts.

The difficulty McGinn (1989/95) faces, however, is it seems clear that upon successfully forming the 'hard' problem, we employ an accurate conceptual formulation of this non-spatial property, which seems to necessarily simultaneously unite our understanding of spatial concepts with our understanding of introspective concepts; indeed, it is precisely this unified conceptual formation that underpins our understanding of the 'hard' problem, and, arguably, makes it so 'hard'. Kriegel (2003) seems to acknowledge this upon asking: 'How could we formulate the problem without employing the missing concept?' (p. 186). So, as Kriegel typifies, in forming an understanding of the problems pertaining to the concept of property P (the non-spatial property of experiential consciousness), McGinn (1989/95) seems to necessarily presuppose a connected, simultaneous understanding of introspective and perceptual concepts, and, as this conceptual unity seems to act as the prerequisite for both McGinn's arguments and Chalmers's (1996) formulation of the 'hard' problem, it seems McGinn must contend that just as a formulation of the problem of consciousness is not causally closed to us, nor is a potential solution[10].

Thus, whilst it seems clear that such a solution will not magically present itself by employing spatial observation, it is precisely our capacity to understand the special nature of the concept underpinning the problem that proves the efficacy of our 'epistemic tools', and hints at our potential to produce solutions (see chapter 6). As such, far from resigning ourselves to the deficiencies of our epistemic equipment, we must find a way to accommodate such non-spatial equipment into our metaphysical

[10] I note here that McGinn (1989) does consider this objection amidst his formulation of 'Humean minds' (p. 365). However, for my purposes in this paper, I do not deem it necessary to devote any more time to McGinn. Instead, I maintain that the formulation of the 'hard' problem and its potential solutions explicated herein are themselves evidence for the contention that we *do* hold the concept-forming capacities required to understand the non-spatial property 'P'.

framework. As, without this accommodation, we bring the entirety of both our metaphysical, and, as shall be explored in chapter 2, epistemic presuppositions into question.

Chapter 2

A further problem: epistemic justification without phenomenal experience

That all our knowledge begins with experience there can be no doubt. For how is it possible that the faculty of cognition should be awakened into exercise otherwise than by means of objects which affect our senses, and partly of themselves produce representation … to compare, to connect, or to separate these, and so to convert the raw material of our sensuous impressions into a knowledge of objects, which is called experience? In respect of time, therefore, no knowledge of ours is antecedent to experience, but begins with it.

(Kant 1781, p. 37)

Picture the scenario outlined within the introductory paragraph of this thesis: I, this subject of experience, softly indent the keyboard and play witness to the letters formed upon the screen. Subsequently, and by virtue of repeated phenomenal experiences of this kind, I gain the knowledge (defined as justified belief[11]) over the causal interaction between the screen, the keyboard and myself that allows me to efficaciously interact with my surroundings so as to produce this thesis. This scenario typifies the causal relationship between experience and knowledge that is widely held by non-philosophical communities. Firstly, light refracts off the screen and keyboard, before entering my cornea to produce a phenomenal, visual experience. From here, I place my hand upon that which I see, and my somatosensory neurons respond to the external keyboard, creating the phenomenal sensation of 'touch' that causally links

[11] This depiction is in line Plato's (380BC) definition, and typifies the widely held interpretation of knowledge as justified true belief.

me to my surroundings. Finally, my visual experience of the letters upon the screen changes upon my having this sensation of touch, and, as a result, I form a belief pertaining to the causal relationship between the screen and my sensory, phenomenal experience of touching the keyboard. Thus, we present a formation that seems highly intuitive: we hold phenomenal, 'what it is like' experiences that, over time, provide justification for our beliefs and, eventually, confer knowledge.

The issue, however, is that if the contentions within chapter 1 are upheld, this intuitive relationship between experience and knowledge acquisition presents a damning problem for physicalism, as the phenomenal experience that acts as our most fundamental epistemic tool (what Sellars 1956 terms the 'given', p. 128) and forms the foundation for all our epistemic justifications, is not explained by the spatial observations employed by physicalist metaphysics. Thus, if this problem holds, the entirety of the physicalist epistemic framework collapses, by virtue of failing to explain the non-spatial, phenomenally experiential property that predicates and justifies all physical knowledge.

Historically, philosophers have attempted to save physicalism from this charge of epistemic vacuity (see Ryle 1949, Rorty 1979, Sellars 1963 and Davidson 1986) by denying the relevance of this purported 'problem' altogether; instead upholding the contention that phenomenal experience is distinct from the concepts and beliefs that inform our understanding of the world, and, as such, holds no relevance to epistemology. As a result, Chalmers (1995) attempted to distinguish the 'easy' problems of explaining 'the integration of information by a cognitive system' and our 'ability to discriminate, categorise, and react to environment stimuli' (p. 201) (or the problems of how we come to know the world, henceforth referenced as the problem

of epistemic justification), from the 'hard' problem of explaining our 'what it is like' experience. Thus, philosophers commonly maintain that the problem of epistemic justification is potentially explainable using physicalist metaphysics, and, in line with Chalmers (1995), tend to uphold the divorce of this 'easy' problem from the 'hard' problem of experience.

In this chapter, I attempt to explicate the underlying contentions implicit to philosophers of this kind, before arguing in favour of the contemporary movement (see Brewer 1999, McDowell 1996 and Pryor 2000, Smithies 2014), which has arisen as an attempt to highlight the relationship between epistemic justification and phenomenal consciousness, and close the divide between this purportedly 'easy' problem and our 'hard' problem of experience. From here, I highlight that, if this contemporary movement is correct, physicalist metaphysics must concede that phenomenal experience plays an implicit role within our epistemic justifications, and, as such, physicalism must face a further epistemic issue that any purported 'solution' to the hard problem must address.

By means of explication for the thesis that denies the link between experience and epistemic justification, let us return to the depiction of knowledge acquisition outlined at the beginning of this section, in which my phenomenal experience of the causal interaction between myself, the keyboard and the screen acts as justification for all further beliefs related to my interactions with these environmental stimuli. The problem with this, as Davidson (1986) highlights, is:

'The relation between a sensation and a belief cannot be logical, since sensations are not beliefs or other propositional attitudes. What then is the relation? The answer is, I

think, obvious: the relation is causal. Sensations cause some beliefs and in this sense are the basis or ground of those beliefs. But a causal explanation of a belief does not show how or why the belief is justified.' (p. 229)

Thus, whilst it seems judicious to argue that my belief in the existence of this keyboard is justified by the experience inherent within my subjective 'what it is like' visual sensation that occurs upon light refracting from this object, Davidson (1986) highlights that whilst my phenomenal experiences may indeed prove to be the causal grounding for my beliefs about keyboards, we face a difficulty upon attempting to glean objective justification for such beliefs whilst employing experience alone. This line of argument seems to typify the philosophical contentions underpinning all those who appeal to the divide between the phenomenal experiences 'which are presented or given to the mind', and the epistemic 'constructions or justifications' which occur as a consequence (Lewis 1929, p. 52), and it is this divide that acts as a grounding for what Sellars (1963) terms the 'myth of the given'.

Thus, Sellars upholds Lewis's (1929) appeal to divorce experiential sense data from the justifications I bring to this data, contending that 'all awareness of sorts, resemblances, facts…all awareness of abstract entities – indeed, all awareness even of particulars – is a linguistic affair' (Sellars 1963, p. 29). Meaning, one can only be justified in a belief about 'greenness', for example, if one already holds the conceptual and linguistic knowledge pertaining to 'greenness', as 'the concept of looking green, the ability to recognize that something looks green, presupposes the concept of being green' (Sellars 1963, p. 146). Hence, the purported 'givenness' of our sense data is indeed a myth, according to Sellars, as any knowledge inferred by

experience is not justified, or 'given', by virtue purely of itself, but instead occurs only as a result of the conceptual justifications that reasoning instils upon this underlying data. Thus, for Sellars, my experience of this keyboard is not evidence of the keyboard in and of itself; instead, I must first have learnt the concepts underlying the object in order to notice that the object in front of me is indeed a keyboard. As such, sense data is an insufficient means to glean knowledge, as 'instead of coming to have a concept of something because we have noticed that sort of thing, to have the ability to notice a sort of thing is already to have the concept of that sort of thing and cannot account for it' (Sellars 1963, p. 176). Simply, then, Sellars' argument presents a polemic against the mental realists and empiricists, who had attempted to posit experience as the 'given' that acts as our 'epistemic bedrock' (McGrew 2007, p. 57), and is predicated upon an attempt to deny the notion that knowledge is rooted within experience, via upholding the central claim that we must divorce the sense data inherent within experiential states, from the state of justified 'knowledge' that occurs upon performing the conceptualisations that place this sense data into the 'logical space of reasons' (Sellars 1963, p. 36).

The foundation of Sellars' (1963) contentions can be deconstructed into two opposing arguments of varying intensity: the former is a weaker attack on empiricism, which is predicated upon the notion that experience is *entirely* superfluous to the process of epistemologically justifying inferences to the external world, whilst the latter maintains that, even if experience *is* in some way implicit to the process of epistemic justification, it can not be justified without appeal to concepts and, therefore, cannot act as the 'given'. In what follows, I present attacks on both arguments, before concluding that, ultimately, even Sellars' (1963) stronger argument must, at present,

rely upon phenomenal experience to explain our capacity for concept formation, and, as such, I maintain that Sellars cannot easily substantiate his claim that phenomenal experience is not, in some sense, the 'given'. I begin with an elucidation of the problem associated with Sellars' 'weaker' attack on empiricism.

The most immediate problem with this weaker argument, which attempts to entirely disentwine experience from the process of epistemic justification, is that, in any conceivable account of knowledge acquisition, the perceptual experience associated with sense data remains inherent to the process of achieving justification for our inferences to an external world. This sentiment seems to be shared by McDowell (1996), who argues that if we did entirely renounce empiricism as Sellars (1963) and Davidson (1986) suggest we must, our justification for beliefs or inferences to the objective world become epistemologically 'blind' (p. 66), by virtue of removing the empiricism which accounts for our relationship to the world as an existent 'actuality' independent of thought, and resigning ourselves to, what McDowell (1996) terms, a 'frictionless' (p. 66) stream of unjustified thought that foregoes the distinction between thought and world. Thus rendering redundant both any attempt to justify thoughts pertaining to the external world, and any attempt to justify a belief in the content of thought A over the content of thought B. Here, McDowell (1996) is employing the Kantian (1781/87) sentiment that 'thoughts without intuitions are empty; intuitions without thoughts are blind' (A51/B75), and highlights that, if the prima facie, non-conceptual, intuitive content derived from the perceptual experiences that account for our 'immediate relation…to objects' (Kant 1781/87, A19/B34) is removed, our thoughts are stripped not just of the content required to justify inferences to an external world, but are stripped of all content, so as to become a

stream of ‘empty thoughts’ that exist as nothing beyond ‘frictionless spinning in a void’ (McDowell 1996, p. 11).

Thus, McDowell (1996) highlights that if experience is entirely removed from the process of epistemic justification, we lose the content that acts as the epistemic ‘friction’ required to transition our inferences from a potential perpetual stream of unjustified, contentless beliefs about how the world may be, to justified, content-rich, beliefs about how the world is in actuality. As such, McDowell (1996) convincingly argues that experience must be, at least partially, implicit to the process of epistemic justification, and, in line with Crane (2013), maintains that ‘it is not clear what remains of the Sellarsian attack on the given’ (Crane 2013, p. 232), as justification is only gleaned upon employing our capacity for perceptual experience so as to bring ‘concepts to bear on what you see’ (Crane 2013, p. 232).

Whilst McDowell’s (1996) and Crane’s (2013) arguments are simply employed to highlight that the epistemic justification for beliefs, or inferences, pertaining to the external world is not found within a contentless, frictionless relation between various concepts or beliefs, but is instead found within the relation of these concepts to the epistemic friction that arises from content-rich perceptual experiences. My proposal, in line with Smithies (2014) and Pryor (2000), is that if concepts cannot be justified without plotting said concepts alongside that which one sees in a perceptual experience, and one cannot gain knowledge over what one sees without first having the phenomenal experience (See Jackson 2003), then we may infer that the phenomenal properties that constitute this experience are themselves, at least partially, responsible for the process of forming justified beliefs about the external world. As

such, I argue that regardless of whether one agrees with McDowell (2008) & Pryor (2000), who argue that unified, experiential sense data provides immediate justification for beliefs, or Sellars (1963) & Davidson (1987) who argue that sense data only provides justification upon being brought into the realms of reason, one must either address the issue of epistemic 'blankness' that arises from attempting to establish justification without employing *any* experiential content (as highlighted by McDowell 1996 and Kant 1781/87), or accept that phenomenal experience is, at least partially, implicit to the process of epistemic justification.

Whilst this line of argument seems difficult to refute, and is arguably successful in showing that the experience implicit to our having sense data is in some way equally implicit to epistemic justification (and for the purposes of this thesis this claim is arguably enough, as this alone seems to bind the easy problem of epistemic justification to the hard problem of experience), it does not, however, refute Sellars' (1963) stronger claim pertaining to the non-inferential nature of experience. This stronger argument does not fall foul to the problems of the weaker argument, and instead maintains that whilst experience *may* play a role in epistemic justification, it is not immediately justified, or a 'given', as, ultimately, it is still nothing beyond sense data that must be moulded and brought in line with reason in order to be justified. This line of argument proves beneficial to anti-realists about the mental, as it maintains that an explanation of the 'given' involves the easy problem of explaining our capacity for conceptualisation and reason, and thus distances our 'epistemic bedrock', or Sellarsian 'given', from the hard problem of explaining phenomenal experience.

This, however, fails to convincingly avoid similar issues to those erected against Sellars' weaker claim that experience plays no role whatsoever in justifying beliefs. As, if we uphold the intuitive, and widely endorsed, notion that concept formation consists in the 'ability to compare what is represented in one experience with what is represented in others' (Smith 2016, p. 88), then, as I shall endeavour to argue, the currently championed theories of physically constituted concept acquisition must inevitably confront the problem of accounting for how concepts are initially formed without relying upon forms of innatism or externalism, and, in so doing, must either (potentially) erroneously place absolute faith in the some future scientific breakthrough to explain how concepts are initially formed non-experientially, or rely upon phenomenal experience to solve this 'concept grounding problem' (Dorrfner and Prem 1993). If this can be shown, I argue that as, at present, our best solution to this grounding problem is found upon positing phenomenal properties, we cannot easily divorce the 'hard' problem from the 'easy' problems of 'the integration of information by a cognitive system', or our 'ability to discriminate, categorise, and react to environment stimuli' (Chalmers 1995, p. 201), and maintain that, without an adequately unified physical solution to these problems, Sellars potentially remains reliant upon phenomenal properties to ground the concepts employed within his process of epistemic justification, and therefore cannot easily substantiate the claim that phenomenal experience is not, in some sense, the 'given'. In order to establish this, I begin with an outline of Papineau's (1993/2002) theory of concept acquisition, before using this as a platform from which to ground a phenomenally constituted theory of concept acquisition that advances an alternative model with the potential to reveal the weaknesses in Papineau's theory, Sellars's stronger argument, and broader theories of physically constituted concept acquisition.

In order to elucidate the deficiencies of Papineau's theory of concept acquisition, it is prudent to first delineate the conditions a successful theory of concept acquisition must avoid in order to not 'beg the question' by presupposing the existence of the very thing it is attempting to explain. Floridi (2012) demarcates these two conditions as follows:

a. 'No form of innatism is allowed; no semantic [or conceptual] resources (some virtus semantica) should be magically presupposed as already pre-installed; and
b. No form of externalism is allowed; no semantic [or conceptual] resources should be uploaded form the 'outside' by some deus ex machine already semantically proficient.' (Floridi 2012, p. 137)

With these conditions in mind, we are in a position to elucidate Papineau's (1993) theory of concept acquisition. This theory maintains that, during the act of forming concepts, we are attentive to a given stimuli, and a purely physical mechanism in our brain occupies a relation to both present and past iterations of this particular stimuli 'wherein incoming stimuli are compared with some stored pattern, and a match between them is registered' (Papineau 1993, p. 120). In this model, our capacity to form concepts is constituted upon a disposition to form and recognise relations between disparate experiential, neural, or perceptual states. Papineau (1993) argues that this relational power is entirely reducible to our physical brain's capacity to recognise distinct, or similar, neural configurations, and maintains that, in the case of Jackson's (2003) Mary argument (see section 1.2 of this thesis), in which the subject,

Mary, experiences 'redness' for the first time, 'Mary simply acquires a [neurologically constituted] "non-conceptual template" which can be compared directly to further experiences so as to cause Mary to believe that she is experiencing red again' (Papineau 1993, p. 110). Papineau (1993) is careful to note here that 'she (Mary) doesn't arrive at this belief by noting that the experience has property P, and concluding that is an experience of seeing red. [Instead] there is simply a [purely physical] mechanism in her brain which compares the experience with the template which yields this belief directly' (p. 110). So, we have a model of concept formation that rests upon, what Papineau (2002) terms, the 'underlying power of perceiving *as*' (p. 108), which can be exercised so as to 'form concepts which enter into fully fledged judgments' (p. 108), but remains, in itself 'perceptual rather than judgmental' (p. 108), and wholly reliant upon the brain's capacity to store templates of a perceptual experience so that our neural patterns 'resonate with incoming signals' (Papineau 2002, p. 120).

With this framework articulated, we witness that the immediate problem with this particular theory of concept acquisition is that, as Floridi (2012) articulates, it seems to face the hurdle of establishing precisely how a neurological mechanism may pick out a neurological pattern, so as to form a concept, without first relying upon a form of innatism to gift the brain with a pre-installed conceptual resource. As, if Papineau's brain mechanism is capable of producing concepts by matching, or interpreting, a neurological pattern 'red' to be one and the same as a previous neurological pattern, we are left asking precisely *how* this initial capacity to match neurological pattern red 1 (NPR1) to neurological pattern red 2 (NPR2) occurred without the brain already holding the concept of red? Indeed, if, as Papineau's formation demands, the

neurological pattern for red is non-conceptual initially, and the brain notices a repeated pattern so as to pick this referent out as 'red', then, in this process of noticing, matching, or interpreting, Papineau seems to grant the brain a pre-encoded concept of redness that is employed within this process of matching, or interpreting[12]. Simply, in Papineau's account of concept formation, we are left with no adequate explanation for how concepts are initially formed, and worse, the explanatory success of his framework seems reliant upon the presupposition of the very thing he is attempting to explain.

This reliance upon an innate, pre-installed conceptual resource to account for the brain's capacity to conceptualise is not unique to Papineau's theory, and arguably afflicts all theories of purely physically constituted concept acquisition, in which the purely physical-functional properties of a neural mechanism are said to account, in entirety, for our capacity to compare our initial neural configuration for red alongside subsequent red neural configurations, in a manner that produces concepts by 'picking out', or 'matching/identifying two or more subsequent patterns with the original one' (Smith 2016, p. 86). This is a problem, because in order to establish how such physical properties achieved this disposition to 'pick out' distinct referential types, it seems that the neural mechanism must either be pre-installed with a conceptual resource (and thus face Papineau's issue of violating one of Floridi's conditions), or be equipped with the disposition to initially form concepts by non-conceptually, and non-experientially, 'picking out' the informational differences between, for example, the content of referent 'red' (R), and the content of referents ~ R in a manner that produces new conceptual information about the nature of these referents. Yet, in this

[12] Smith (2016) employs a similar objection against Papineau (1993).

latter case, the physicalist is left with no adequate explanation for what this neural mechanism might look like, and, as such, remains reliant upon some future scientific breakthrough to account for both precisely where in the brain the neural mechanism that is capable of unifying the informational content of R and ~ R is located, and, further, precisely how this unifying neural mechanism is capable of non-experientially achieving the disposition to 'pick out' the informational disparities between competing referents so as to produce concepts (this particular problem is referred to by Floridi (2012) as the 'symbol grounding problem')[13].

Indeed, whilst various physicalist theories of concept acquisition have attempted to delineate the nature of this unifying neural mechanism (see Schnur et al 2009, Garagnani et al 2008, Freedman et al 2001), no theory has been able to advance a convincing solution to the problems of how this neural mechanism unifies the informational content of disparate referents, and how this neural mechanism achieves the disposition to pick out, or 'ground', competing, or similar, informational content[14]. As such, even if we were to accept that perhaps one of the physicalist theories of concept formation *has* articulated the physical locus in which distinct neural patterns, and in turn informational content, are unified, we would still be left in search of an

[13] Floridi (2012) demarcates this as the problem of accounting for 'precisely how a system can autonomously elaborate its own semantics for the symbols (data) it manipulates and do so *from scratch*, by interacting with its environment and other formal symbol systems' (p. 136).

[14] The most promising areas of research in this area maintain that the unifying neural pattern may be found in the anterior temporal lobe (Garagnani et al 2008), or the prefrontal cortex (Schnur et al 2009), or, more broadly, temporal lobe structures with highly selective responses to objects (Freedman et al 2001). However, all such areas offer competing, but equally vague, solutions to a problem that, at present, seems to over-exert the explanatory power of physicalism. Indeed, whilst these physicalist solutions rest upon the assertion that *perhaps* we may, at some point, infer that these brain regions are, in some way, implicit to concept formation, we are, at present, still left in search of an explanation for precisely how these neural patterns acquire the disposition to distinguish between two or more referential types, and remain in need of an explanation for why, as Kiefer & Pulvermuller (2012) attest, 'several lines of evidence clearly indicate experience-dependent formation of cell assemblies in sensory and motor areas, which code conceptual features in a modality-specific fashion' (p. 816).

explanation for how the purely physical properties of this neural mechanism achieved the disposition to non-experientially 'pick out' the differences, or similarities, between this informational content, in a manner that produces new, conceptual information.

Many philosophers maintain that this particular problem shall be short-lived, as a 'computational', or purely physical-functional explanation will eventually be produced. However, as Searle's (1980) Chinese room argument articulates, both a functional computer and a human subject may theoretically employ a computational programme to 'pick out' symbols in a manner that passes a Turing test being held in the Chinese language, without the human subject holding any conceptual understanding, or semantic information, over the symbols being picked out. Thus, as the human subject is not knowingly 'picking out' the disparities between these symbols in a manner that may lead to, or infer, a conceptual understanding, we may surmise that the physical-functional computer is itself not knowingly 'picking out' such disparities in a manner that may lead to, or infer, a conceptual understanding. As, if it were, then this conceptual understanding would be contained explicitly within the same computational programme employed by the human subject, and, the human subject would hold an understanding of such concepts that would allow them to knowingly 'pick out' the differences and/or similarities between the symbols. As a result, Searle (1980 maintains that (1) concepts are not grounded externally, and (2) the purely physical-functional computations of the brain are not enough to meaningfully ground concepts, and, as such, there must be some additional property, or mechanism, in the brain that allows us to *initially* acquire, or form, concepts (Harnad 2001). Indeed, as Harnaad (1990) articulates, the fundamental problem is one

of accounting for how the semantic, or conceptual, content of a formal symbol system be made intrinsic to the system, 'rather than just parasitic on the meanings in our heads' (p. 336), or, more explicitly, how are concepts initially be formed by a brain that, if Searle's thought experiment is correct, cannot necessarily rely upon purely physical-functional computations to 'pick out' the informational differences that initially lead to the formation of meaningful, conceptual information[15].

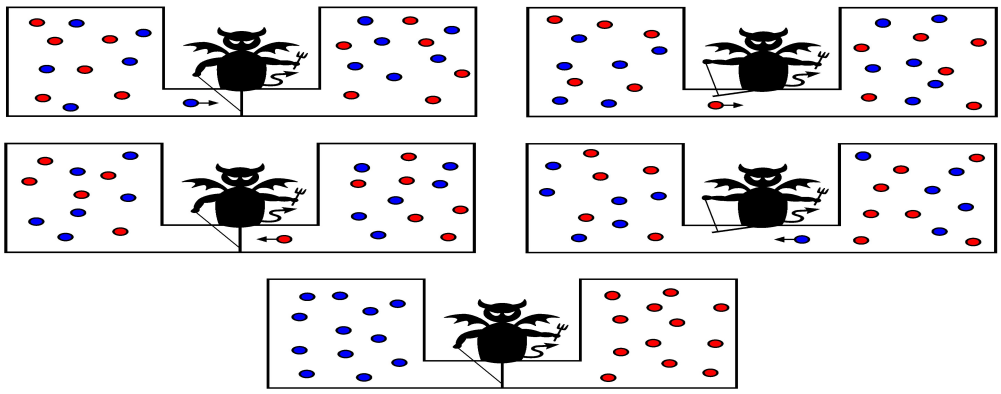
The best current theory of how this new, conceptual information may be produced, or 'grounded', non-experientially, is found within Floridi (2012), and asserts that the informational content inherent within machines 1 & 2 potentially becomes new, semantic information when integrated within an additional functional machine (see Floridi 2012, p. 169-172). However, this is not a convincing solution to the 'concept grounding problem' for physicalism, as if we take the information contained within the neural patterns 1 & 2, and then add further information contained within an additional neural mechanism, it seems we have acquired no new conceptual information, and instead have simply achieved a conglomeration of the pre-existing informational content contained within the initial neural patterns '1', '2', and the neural mechanism '3'. Yet, as established, this is not enough for concept acquisition, as the neural mechanism must not simply add to the total informational content by acting as a locus for information, but instead must be able to reduce the potential entropy of this information by unifying it in such a manner that new information is produced by virtue of delineating the differences, or similarities, between the informational content of '1' and '2'. In this regard, it is not enough to simply add an

[15] Horgan (2013) furthers this notion by highlighting that 'the real moral of Searle's Chinese room thought experiment is that genuine original intentionality requires the presence of internal states with intrinsic phenomenal character that is inherently intentional' (p. 233).

extra neural pattern, or an 'Artificial Agent' (AA) of the type employed by Floridi (2012, p. 170), so as to simply amalgamate the informational content of neural patterns '1', '2' within neural mechanism '3', but instead we must imbue the neural mechanism, or AA, with a further disposition that allows it to act as a unifying locus 'X', which is capable of both binding the potentially disparate informational content 1, 2 in such a way that the content is unified so as to become 1X, 2X, and then non-computationally 'picking out' the differences, or similarities, between the informational content of 1X and 2X when unified in this manner. However, as Searle's (1980) argument highlights, and Floridi (2012) himself notes, this disposition to 'learn how to use associated symbols [informational data] or internal states' (p. 173) so as to 'pick out' the differences to produce concepts is difficult to account for physically, and, as such, we witness frameworks, such as Papineau's (1993-2002), which seem to presuppose a capacity for concept formation, and, in so doing, invariably 'beg the question' (Floridi 2012, p. 173). Indeed, as articulated, there is currently no convincing articulation for how this neural mechanism might achieve the disposition necessary to unify the informational content in this way, and, as such, we are left turning to a phenomenally constituted notion of concept acquisition to explain how we initially produced the information necessary to form the concepts employed by physicalist metaphysicians.

This phenomenal, or experience-dependent (see Kiefer & Pulvermuller 2012), model of concept acquisition may best be articulated in terms of an experientially constituted

'Maxwell's Demon'[16], which may be most eloquently articulated by the use of the following diagram:



As can be inferred, in the initial diagram (top left), the demon has an experience of the informational content inherent within blue particles, and stores the experience of this informational content within memory, or, what Papineau (1993) terms, a neural pattern. From here (top right), the demon experiences the informational content inherent within red particles, and the same process of neural storage occurs. At this stage, the demon is able to employ his experiential token, or 'stored pattern' (Papineau 1993) for red and blue experiences, and 'pick out' the experiential differences between them. Thus, in the middle two diagrams, we witness the demon having perceptual experiences, instantly demarcating these experiences as either an experience of 'red' or 'blue', and, finally, in the bottom diagram, we witness the demon reducing the informational entropy of the red and blue particles by segregating them into two distinct experiential types. All of this occurs by virtue of the demon's capacity to extract the information that occurs between 'the gaps', or differences, in red and blue experiential types. Meaning that, at foundation, the demon works by

[16] Maxwell's demon was initially constructed by Maxwell (1871) as a thought experiment in the philosophy of physics, and was designed to reveal how the second law of thermodynamics may potentially be violated by the integration of an intelligent agent.

virtue of understanding what it means to experience red as distinct from what it means to experience blue, and so, at this stage, we may infer that the demon has a perceptual understanding of what red and blue are, which, under Papineau's model of concept acquisition, could be 'exercised… to form concepts which enter into fully fledged judgments' (Papineau 1993, p. 108).

Thus, using this model, we highlight that the only implicit difference between a phenomenal, or experiential, model of concept formation, and a physicalist theory of the type articulated in Papineau's model, is that Papineau places absolute impetus upon the purely physical, functional properties of the demon's brain to accommodate for the power to delineate, or pick out, distinct neural, or perceptual types, and, in turn, produce concepts (and faces the inevitable difficulty of accounting for this power without presupposing concepts). Where as the phenomenal, or 'experience-dependent' (Kiefer & Pulvermuller 2012), model asserts that the demon's brain holds phenomenal, experiential properties, which subsume into a conjoint phenomenology for a singular experiential subject, so as to imbue the demon with the disposition to delineate between experiential types, and, in turn, the disposition to encode new neural patterns that may represent the information necessary for concepts. In this latter model, upon experiencing the red particles, the experiential type is mapped to the demon's memory as a neurologically constituted pattern that denotes a token experience of the red experiential type, and the demon experientially demarcates incoming particles as holding an experiential type that matches, or does not match, his memory of the experiential type of red particles. Enabling the demon to contrast the differences in his experiences of red and blue, and extract this information so as to form a new neurological pattern, which contains some semantic, conceptual

information about the properties of red/blue particles, and facilitates the process of experiential segregation. Conversely, in physicalist models such as Papineau's, upon interacting with the particles, the brain encodes a neural pattern that denotes the distinct informational content of red or blue particle, and is subsequently (somehow) able to match incoming stimuli to a particular pattern so as to segregate the particles.

As a result of the implicit difficulties involved in explaining precisely how the purely physical-functional brain achieves this capacity to match, or 'pick out' similarities between neural patterns and incoming stimuli, it should come as no surprise that Papineau presupposes concepts in order to avoid the inevitable difficulty of accounting for how the informational disparities that give rise to concepts may be recognised by the non-experiential brain. Indeed, it seems that the central problem with all physicalist accounts of concept formation, as Papineau's framework exemplifies, is that they leave little room for delineating a dispositional locus from which differences between referents may be realised, and so face the inevitable issue of accounting for how the easy problems of 'the integration of information by a cognitive system' (Chalmers 1995, p. 201), or our ability to discriminate between the informational differences necessary to form concepts, may be solved within a model of a purely physical, functional brain that cannot, at present, account for the neurologically unified mechanism that initially picks out, or produces, the information necessary for concepts. In order to achieve this solution, it seems we remain reliant upon positing phenomenal properties as implicit experiential qualities of the brain's neurological patterns, and, from here, establishing that such 'experiential parts' subsume so as to constitute our experientially unified phenomenal states, in which the token, or type, experience of red is experientially evident to the same phenomenally

experiential subject as the type, or token, experience of blue, and the concept of red arises as a natural corollary of the informational content inherent within the differences between these distinct experiential types when presented to one phenomenally experiential subject.

At this stage, the physicalist may suggest that this phenomenal mode of concept formation is itself reliant upon an explanation for how phenomenal properties may produce experiential unity. To this, in line with Chalmers & Bayne (2003), I would argue that if phenomenal properties inhere within the physical brain in such a way that each neural pattern is an experiential part of a wider whole, then, upon these experiences occurring simultaneously within one consciousness, they become subsumed into a wider experiential whole that exemplifies 'a conjoint phenomenology for both states' (Chalmers & Bayne 2003, p. 37). Such that, for example, upon the neurological pattern for 'blueness' occurring simultaneously with the neurological pattern for 'redness', these experiential parts will be unified and subsumed into one experiential subject, for whom "there is something it is like to be in two states simultaneously" (Chalmers & Bayne 2003, p 32). Whilst this in itself is not a wholly convincing articulation of how the unified subject of experience arises initially, I argue that this particular explanatory gap (what James 1896 terms the culmination problem) is not enough to save the physicalist from the difficulties faced by a purely physical explanation for concept formation. As a result, I maintain that the physicalist must cease to assume that the 'hard' and 'easy' problems are wholly distinct, and instead must conclude that because our best current explanations for how conceptual frameworks are formed are experience-dependent, a physical explanation for the 'easy' problems must simultaneously address the 'hard' problem by explaining away

our seeming reliance upon phenomenal experience to ground (and justify) concepts, whilst non-computationally and non-experientially accounting for the disposition to 'pick out' the informational differences that are necessary to form the conceptualisations that hold our epistemic framework together. Without this explanation, the physicalist strategies that attempt to entirely disentwine phenomenal experience from our epistemic bedrock must accommodate for the possibility that a purely physical solution to the symbol grounding problem will not be articulated, and, in so doing, must address the fact that a solution predicated upon the existence of phenomenal properties is simply far more parsimonious than a solution predicated upon the non-existence of such properties.

As such, I have highlighted that Sellars' weaker argument must address the issue of accounting for the epistemic 'blankness' that arises upon denouncing empiricism entirely, and, from here, have maintained that those who endorse Sellars' stronger argument, which attempts to entirely disentwine the 'hard' problem of explaining experience from the 'easy' problems (Chalmers 1995, p. 201), must first provide a fitting non-experiential solution for these easy problems, which is capable of explaining away our seeming reliance upon phenomenally constituted experiential unity to ground the concepts employed within epistemic justifications.

2.1 The fundamental problem and three potential physicalist solutions

In sum, even if the 'symbol grounding problem' is one day convincingly physically solved, it remains difficult to refute that experiential sense data is, even if only partially, implicit to the process of epistemic justification, as the accuracy of a

concept is only known by virtue of its relationship with sense data. Thus, the physicalist cannot avoid the problem posed at the beginning of this chapter, as a failure to physically solve the 'hard' problem would result in epistemic justification being gleaned from a non-physical experiential property that is both inexplicable, and entirely inconsistent with, the monism of physicalist metaphysics. The fundamentals of this issue can be exemplified as follows: If we cannot account for sense data without appealing to phenomenal experience (Q), and sense data is implicit to the epistemic justifications that act as the foundation for our understanding of physicalism (P), then physicalism must be either A) self-refuting by virtue of allowing for this non-physical experiential property, or B) epistemologically vacuous by virtue of gleaning justification from a non-physical property (Q) that is incommensurable to the physical quantifications underpinning the physicalist epistemic framework.

So, if physicalism denies the unified, experiential identity 'Q', it loses the property from which it gleans epistemic justification, yet if it accepts that Q exists as an experiential identity, it becomes self-refuting by virtue of allowing for a non-physical property - unless, of course, it can explain that experiential properties are physical. The issue becomes that: as it can be shown that the experiential identity Q is required in order to justify knowledge, and, therefore, that Q acts as justification for the observations and quantifications that underpin physicalism (P), then physicalism risks being shown to be epistemologically bankrupt by virtue of gleaning epistemic justification from an experiential property that acts as the physically incommensurable antithesis to a monistic physicalist metaphysic that can only affirm the existence of quantifiable, spatial structures or functions. As a result, it seems, the physicalist is left with little choice but to accept that Q is, in some way, implicit to the epistemic

justifications that underpin physicalism (P), and, further, as Q seems to present an experiential identity that is beyond reconciliation with P, it seems the physicalist must also accept that the implicit link between epistemic justification and the experiential identity of Q risks collapsing the entirety of the physicalist epistemic framework into the realms of epistemic vacuity, due to P gleaning epistemic justification only by virtue of the non-physical, experiential identity, which, according to physicalism, does not exist. Meaning that the hard problem cannot be divorced from the purportedly 'easy' problem of epistemic justification, if, as I have attempted to argue, the seemingly physically unquantifiable phenomenal identity of Q is, even if only partially, implicit to epistemic justification.

The ontological gap that underpins this argument seems truly damning to physicalism, and, as a result, the most efficacious physicalist solutions attempt to rectify the property that lies at the heart of the hard problem via denying the existence of Q as an experiential, non-physical identity, whilst simultaneously providing a physical framework capable of explaining our capacity for knowledge. The physicalist 'solutions' that purportedly provide an account of this kind begin by either denying the experiential identity of Q entirely (Dennett 1991), positing a means from which to massage the identity of Q so as to reduce our phenomenal experiences to the brains veridical representation of external content (Tye 2000, Dretske 1996), or contending that the experiential identity of Q is, at foundation, reducible to a conceptual misunderstanding that occurs upon the brain trying to reference itself (Loar 1990/9, Balog 2009/12). It is these attempts to massage or deny the troublesome identity of Q that exemplify physicalism's best means of maintaining its metaphysical and epistemic frameworks, and, as such, the following three chapters shall be devoted to an examination of these physicalist solutions.

Thus, within chapters 1 and 2, I have highlighted that the epistemic gap presented by the hard problem is predicated upon a deeper ontological gap, between the identities of consciousness and physical substances, which presents a deep issue for a physicalist metaphysic predicated upon spatial observations. From here, I have explicated how this ontological gap leads to an epistemic gap that, if not accommodated for, produces a vacuity that is not applicable solely to consciousness, but instead, potentially tarnishes the entirety of the physicalist epistemic framework. Finally, I have highlighted the three most efficacious physicalist solutions to the hard problem. The first of which, Dennett's (1991) illusionism, shall be explored in the subsequent chapter.

Chapter 3

Dennett's Illusionism

Dennett's (1991-2017) illusionism represents an attempt to dissolve the ontological gap underpinning the hard problem, and in turn dissolve the problems inherent within gleaning epistemic justification from a non-spatial property, via eliminating the existence of non-physical, experiential properties entirely. In line with other illusionists (see Humphrey 2011, Frankish 2016, Blackmore 2003), Dennett employs a similar notion to the Husserlian 'natural attitude' (outlined in the introduction of this thesis), claiming that the objective quantifications employed by physicalism represent what is reliable and trustworthy (Dennett 1991 p. 85), and, as a result, any judgements we make pertaining to phenomenal experiences that are incommensurable to the physical sciences must be identified as both illusory and epistemologically unreliable. Thus, Dennett's fundamental aim is to employ a 'heterophenomenological' method, which he champions as 'the scientific method applied to the phenomena of consciousness' (2001, section. 1, para. 8), in an attempt to posit that: if first-person reports of phenomenal consciousness are shown to be physically unquantifiable, we must conclude that such accounts are expressing nothing beyond the 'mistaken belief' (Dennett 1991, p. 85) that phenomenal experiences exist as anything more than an illusion. This chapter aims to erect an attack upon Dennett's claims for illusionism, via highlighting the absurdity of a heterophenomenological method that is underpinned by the very experiential property it is attempting to eliminate, in the hope that, upon debunking the foundation of his enquiry, I may simultaneously remove the foundation from which Dennett posits his eliminativism, and, in so doing, reveal how

Dennett is left incapable of avoiding the same problems erected against Sellars (1963) in chapter 2.

Dennett's heterophenomenological method is touted as the "the bridge between the subjectivity of human consciousness and the natural sciences" (Dennett 2007, p. 249), and attempts to provide a means from which to neutrally and objectively plot the 'real goings-on in people's brain' (Dennett 1991, p. 85) alongside the first-person reports of what is believed to be the 'real goings-on'. Thus, Dennett's heterophenomenology is predicated upon a need to remain 'agnostic' (Piccinni 2010, p. 87) about the epistemic efficacy of first-person reports, so that we may develop a:

'*Cautious, controlled way of taking subjects seriously, as seriously as they could possibly be taken without granting them something akin to papal infallibility, while maintaining (contrary to everyday communicative practise) a deliberate bracketing of the issue of whether what they are saying is literally true, metaphorically true, true under-an-imposed interpretation, or systematically false in a way we must explain.'* (Dennett 2007, p. 252)

Thus, Dennett advances the central claim that, contrary to the Cartesian notion that we indubitably know subjective 'first-person' experiential states prior to knowing anything else, these first-person reports pertaining to 'what it is like' to undergo a given phenomenal experience are not infallible, and, as such, should be taken as nothing beyond raw, informational data that may be plotted alongside the less fallible, scientifically quantifiable goings on within the physical brain, so as to produce an objective, 'third person approach to consciousness' (Dennett 1991, p. 98). Thus,

Dennett employs this heterophenomenological framework as a means to propose that it is only upon mapping these 'fallible' first-person reports to the 'real goings-on' (Dennett 1991, p. 85) within the objectively quantifiable physical brain that we may come to understand the efficacy of an individual's beliefs pertaining to their subjective, experiential states. Further, Dennett (1991) contends that if we find that there is a disparity between these fallible first-person reports and our reliable quantifications of the physical brain, then we would be justified in questioning the reliability of the first-person report. The core of this argument is exemplified as follows:

'If we were to find real goings-on in people's brain that had enough of the 'defining' properties of the items that populate their heterophenomenological worlds (i.e. the subjective worlds projected by what people say in describing their own minds), we could reasonably propose that we had discovered what they were really talking about. And if we discovered that the real goings-on bore only a minor resemblance to the heterophenomenological items, we could reasonably declare that people were just mistaken in the beliefs they expressed.' (Dennett 1991, p. 85)

This argument represents the foundation from which Dennett constructs his illusionism, as the heterophenomenological method purportedly highlights that whilst the raw data inherent within first-person reports of intentional states *does* correspond to the 'real goings-on' in the physically quantifiable brain, and should, as such, be deemed to be scientifically reliable; there is no such correlation to be found within reports of phenomenal states. Thus, Dennett (2016) proposes that, whilst beliefs in physically quantifiable 'intentional objects' (p. 71) are accountable upon mapping

these objects to the representational 'encoding of features of a stimulus' (Frankish 2016, p. 19) in our somatosensory cortex; there is no such means of accounting for phenomenal states. Meaning that whilst my belief that 'I am holding a blue coffee mug' is explicable by virtue of the physically quantifiable existence of both the intentional object and my neurologically traceable representational reaction, so that '[I] am caused to believe in the existence of that mug by the mug itself' (Dennett 2016, p. 72), the first-person belief that I am internally holding a phenomenal 'what it is like' experience, which acts as an intermediary between the coffee mug and myself, cannot be wholly accounted for either neurologically or by the intentional object. Thus, Dennett posits that as such scientifically unreliable first-person reports cannot be brought in line with our more reliable understanding of the physical brain, any notion of a unified, phenomenally experiencing 'I, on the inside' must 'turn out to be something rather like a benign user illusion' (Dennett 1998, p. 357).

Using this purported disparity between the unreliable, first-person evidence for phenomenal experience, and the reliable third-person neurological quantifications in the brain, Dennett attempts to construct an alternative 'multiple drafts model' that he proposes as a solution to, what he terms to be, the unreliable, 'illusory' dualism evident within 'Cartesian Materialism', which posits that there is one unified, experiential seat of consciousness, in which disparate neurological states 'all come together' (Dennett 1991, p. 107). So, at foundation, Dennett's (1991) appeal to his heterophenomenology method acts as a means to deconstruct the ' persuasive imagery of the Cartesian Theater' (p. 107), for which he blames our illusion that we exist as unified selves with phenomenal 'what it is like' experiences, in favour of a physically quantifiable 'multiple drafts model', which is predicated upon an attempt to show that

‘there is no reality of conscious experience independent of the effects of various vehicles of content on subsequent action’ (p. 132). This model purports to eliminate notions of a non-spatial locus of phenomenal experience, and, in order to achieve this, relies upon Dawkins’s (1982) theory of memetics in an attempt to show how ‘human consciousness is itself a (non-local) huge collection of *memes*[17] that can be best understood as the operation of a *Von Neumanesque* virtual machine implemented in the parallel architecture of a brain’ (Dennett 1991, p. 210).

So, in line with his scientism, Dennett falls back on computational and evolutionary theory as a means from which to explain our purported ‘illusion’ of unified phenomenal experience. Contending that somehow the illusion of a locus of experience arises due to self-replicating cultural and evolutionary ‘memes’, which represent informational intentional states that ‘physically reside in the brain’ (Dawkins 1982, p. 109), competing, by virtue of natural selection, ‘for replication by human hosts’ (Blackmore 2003, p. 19). These memes are defined by Dennett (2017) as a:

[17] The concept of ‘memes’ was coined by Dawkins (1982) and is defined as follows: "The gene, the DNA molecule, happens to be the replicating entity that prevails on our own planet. There may be others. If there are, provided certain other conditions are met, they will almost inevitably become the basis for an evolutionary process…I think that a new kind of replicator has recently emerged on this very planet. It is staring us in the face. It is still in its infancy, still drifting clumsily about in its primeval soup, but already it is achieving evolutionary change at a rate that leaves the old gene panting far behind. The new soup is the soup of human culture. We need a name for the new replicator, a noun that conveys the idea of a unit of cultural transmission, or a unit of imitation. 'Mimeme' comes from a suitable Greek root, but I want a monosyllable that sounds a bit like 'gene'. I hope my classicist friends will forgive me if I abbreviate mimeme to meme." (Dawkins 1982, p. 192)

'Kind of way of behaving (roughly) that can be copied, transmitted, remembered, taught, shunned, denounced, brandished, ridiculed, parodied, censored, hallowed.' (p. 206)

Such 'ways of behaving' encompass every instance of cultural imitation one can conceive of, and are passed from one system to the next by virtue of language, which Dennett (2017) signifies as the 'lifeblood of cultural evolution (p. 179), and it is this process of informational dissemination that Dennett (1991-2017) has employed to explain the illusion of phenomenal experience. By means of an analogy for Dennett's thesis, let us consider the software that accompanies the physical hardware of the computer I am employing to write this thesis. The software itself exists as a means to optimise the functionality of the hardware, but remains free to be transferred from one computer to the next, as the computational advances in hardware ensure that all modern computers are optimised to accommodate for advances in software. These contemporary proponents of memetic theory (see Dawkins, Blackmore 2003, Dennett 1991) propose that the physical brain is best understood in relation to the hardware of a computer, with cultural 'memes' best understood as the software that transitions from one physical brain to the next by virtue of evolution's tendency to produce physical brains (hardware) capable of effectively assimilating and disseminating this software. So, cultural memes that are effective at optimising the functionality of the brain are passed from one brain to the next and, over time, the species becomes driven by its capacity to optimise its imitation of these successful memetic structures.

Using this model, Dennett (1991) argues that any notion of a phenomenally unified self is explained away as the user-illusion that accompanies the interaction between

physical, intentional content and the memetic, cultural structures that are most conducive to survival, so that the memes involved in classifying oneself as a unified 'I' survive not because there really is 'something it is like' to be a given a subject, but because the memes involved in this behaviour are evolutionary advantageous. Meaning that, at foundation, this illusion of phenomenal experience is nothing more than an evolutionary optimised process made up of 'thousands of memes, mostly borne by language, but also by wordless "images" and other data structures, taking up residence in an individual brain, shaping its tendencies and thereby turning it into a mind' (Dennett 1991, p. 254).

Whilst the core of this argument is ultimately predicated upon a contemporary evolutionary theory of memetics that is beyond the scope of this thesis, the fundamental notion is that the brain encompasses nothing beyond purely physical, intentional states that self-replicate due to the virtues of their cultural and evolutionary successes. In Dennett's (1991) case, these self-replicating processes are most evident within our language, and it is the tendency for language to self-replicate in the face of natural selection that, Dennett argues, creates the formation of an 'I', that eventually leads to further memetic structures such as 'I think' or 'I want' that unite our memes with external, intentional objects. Further, over time, Dennett argues that these simplistic memes coalesce with intentional objects so as to form even broader, more complex memetic structures that account for the beliefs, desires and judgements, which Blackmore (2003) argues explains our illusion of having a 'selfplex', and Dennett (1991) argues accounts for our illusory experience of having an inner phenomenally experiential 'mind'. Thus, contrary to the Cartesian framework, Dennett (1991) describes 'a world, the subject's heterophenomenological world, in

which are found various objects (intentional objects)…which self-replicate to form a narrative, so that (the self) is a fictional object, and so are the objects described, named, mentioned by the heterophenomenologist' (p. 95). Meaning that, at foundation, we are nothing beyond a culmination of disparate intentional content, underpinned by memetic structures that have coalesced into an illusory, unified fiction we reference as the 'self', so that whilst there certainly seem to be 'what it is like' experiences, these are ultimately accountable upon bringing physical, intentional content in line with the evolutionarily optimised memetic structures that form this illusion of being a unified, phenomenally experiencing 'self'. Thus, Dennett (1991) advances the claim that:

'The self constructed by normal human brains is part of the extended human phenotype: humans brains are born with the default expectation that their environment will contain millions of mostly word-borne memes, which they can automatically 'weave' into a narrative, a coherent sequence, that defines the self.' (Zawidzki 2007, p. 94)

As such, Dennett appeals to the primacy of intentionality. Contending that the memetic structures that underpin intentional content explain conscious experience without relying upon any of the properties that are special 'in the ways qualia have supposed to be special' (Dennett 1993 p. 43). Hence, Dennett employs his heterophenomenological method to dispense of notions relating to a unified, inner, phenomenal 'raw feel', and instead posits an intentional, physical depiction of consciousness as 'a kind of mental content, almost a matter of programming' (Brook & Ross 20002, p. 8). In so doing, Dennett attempts to prove that all illusions of

phenomenal experience are accountable upon bringing 'outer' physically intentional content in line with his notion of self-replicating memetic structures which, over time, form a fictional 'self' that forms the 'centre of our narrative gravity' (Dennett 1992, p. 103). Thus, we see that whilst Dennett's explanation of consciousness provides a relatively robust framework, which is capable of accounting for intentional content and our illusory beliefs that we exist as unified subjects with 'what it is like' experiences, it, ultimately, remains predicated upon his heterophenomenological method, which ambiguously 'explains away' or 'ignores' (Carruthers 2005, p. 247) the hard problem of phenomenal experience, via simply defining qualitative content as nothing beyond an illusion that may be accounted for physically.

So, the success of Dennett's framework is reliant upon the success of this heterophenomenological method, which has allowed him to explain a variety of 'easy' problems whilst avoiding a confrontation with the intractable nature of the truly hard problem. However, akin to modern science, Dennett's success with these 'easy' problems seems predicated upon the principle of 'give us one free miracle, and we'll explain the rest' (Sheldrake 2009, p. 2), as whilst science constructs itself upon the 'free miracle' of something from nothing, Dennett's entire physicalist framework on consciousness seems to be constructed upon the 'free miracle' that we deny the efficacy of our own first-person accounts, and accept the illusory nature of the seemingly indubitable phenomenal experiences we know prior to all else. So, whilst science is postulated upon the free miracle that there just *is* something, Dennett's framework is postulated upon the free miracle that experience just *is* an illusion. However, unlike science's promise to eventually explain this one free miracle, Dennett seems adamant that we must ignore phenomenal consciousness entirely, and

instead focus upon how the memetic structures of the brain coalesced to form this illusion. As Frankish (2016) explains:

'Illusionism replaces the hard problem with the illusion problem — the problem of explaining how the illusion of phenomenality arises and why it is so powerful. This problem is not easy but not impossibly hard either. The method is to form hypotheses about the underlying cognitive mechanisms and their bases in neurophysiology and neuroanatomy, drawing on evidence from across the cognitive sciences.' (p. 37)

So, whilst illusionism seems an attractive prospect, and Dennett's contributions to this area are robust, it seems clear that, prior to addressing the problems of explaining the 'illusion of phenomenality', any movement to replace the hard problem with illusionism must be predicated upon a framework that is capable of providing reasons to accept the problems of illusionism over the problems of phenomenal consciousness. In Dennett's case, these reasons rest entirely upon his heterophenomenological method, and this in itself seems to rest upon his aforementioned 'free miracle' that we forego first person accounts and accept the illusory nature of that which is known with most immediacy. Hence, I posit that if Dennett's 'free miracle' is denied, then the illusionism that Dennett employs to 'replace' the hard problem collapses, and we extricate ourselves from Dennett's eliminativism.

With this in mind, I posit two problems with Dennett's heterophenomenology, the first being Dennett's difficulties in maintaining the internal coherence of a method that simultaneously relies upon and denies the efficacy of first person accounts, and

the second being the issue of this aforementioned 'free miracle'. So, prior to an exploration of Dennett's free miracle, I begin with an examination of his heteophenomenological method, and put forth my contention that the fundamental problem with Dennett's (1991-2017) entire oeuvre on consciousness is one of establishing the epistemic foundation for his method. For, it seems difficult to establish how Dennett (1991) constructed his 'third person', objective enquiry into the phenomenon of consciousness without first employing the 'first person', purportedly 'illusory' phenomenal experiences that he is attempting to deny.

Dennett places impetus in the efficacy of the intentional stance, which purportedly provides grounding from which to supplant the epistemic fragility of first-person reports with the more robust notion of third-person evidence. By employing this, Dennett seems to infer that we solve the problem of 'how to combine the perspective of a particular person inside the world with an objective view of that same world, the person and his viewpoint included' (Nagel 1986, p. 118), via relying upon his heterophenomenological method to provide an objective 'account *of* that point of view which is not itself given *from* that point of view' (Williams 1986, p. 6). The problem with this is that Dennett describes the heterophenomenological method as follows:

'*From the recorded verbal utterances, we get transcripts, from which in turn we devise interpretations of the subject's speech acts, which we thus get to treat as expressions of their beliefs, on all topics. Thus, using the intentional stance, we construct therefrom the subject's heterophenomenological world. We move, that is, from raw data to interpreted data.'* (Dennett, 2001, section. 1, para. 3)

Here, the explanatory and interpretive work is contained within the 'intentional stance', which takes the raw data and infers an interpretation as a consequence. Whilst Dennett seems content to favour third-person interpretations over first-person reports, I argue that, if this is case, Dennett's heterophenomenology must necessarily extricate itself from the purportedly unreliable first-person biases of both those having the experience *and* those doing the interpreting. And contend that, if these are extricated, Dennett (1991) fails to account for all the data pertaining to the phenomena of consciousness, and so cannot hope to establish a truly objective explanation for conscious experience. This seems clear, as if we begin with 'raw' data pertaining to person 1's (P1) judgments and beliefs, and task person 2 (P2) with forming their own 'interpretations' (thus necessarily foregoing Dennett's neutral 'intentional stance' via employing their own judgments to interpret the data[18]) in order to plot P1's beliefs about their phenomenal experience alongside P1's physically quantifiable brain, one must ask: how do we extricate ourselves from the purportedly 'unreliable' first-person reports of P2? As such, it seems in order to maintain the internal coherence of his method and justify his contention that *all* first-person reports are unreliable, Dennett must not only posit that P2's judgments hold more reliability than P1's, but that somehow P2's judgments are not first-person, unreliable judgments at all, and, instead, inexplicably make the leap from an interpretative, subjective judgment inferred from the first-person perspective of the objective 'raw data', to an objective judgment that is itself not afflicted by the biases of being derived from a first-person perspective. Indeed, I argue that if Dennett maintains that those doing the interpreting achieve pure objectivity amidst the act of making inferences and judgments upon raw

[18] Chalmers (2010, p. 54) and Wah (2007, p.6) make similar claims.

data, then he seems to be caught in limbo between an interpretation that is simultaneously derived from both the apparent objectivity of the third-person intentional stance *and* the first-person perspectives and subjective interpretations of those doing the interpreting. If this is the case (and it seems that it must be), in order for Dennett (1991) to maintain the coherence of his purportedly objective, scientifically grounded method, he must extricate his method from the unreliability of the judgments gleaned from the first-person perspectives of *both* P1 and P2, but, in so doing, Dennett seems to necessarily remove the judgments that underpin his heterophenomenological method, and, as such, the method collapses under the weight of the first-person, purportedly unreliable, inferences employed during P2's interpretations of the raw data.

This problem becomes more fundamental, however. This is because the necessary integration of P2's judgment seems to leave Dennett reliant upon P2's 'user-illusion' that he is a unified, 'judger' or 'interpreter', withholding a unified, phenomenal experience. For, if we maintain that P2 must employ a first-person judgment in order to 'interpret' P1's 'raw data', and making a justified 'judgement about what is seen is applying one's concepts to what one sees on the basis of this seeing' (Crane 2013, p. 232), then one cannot logically exclude experience from the process of justifying one's judgement. Thus, whilst the heterophenomenological method is reliant upon the judgments of P2, P2's capacity to judge is reliant, at foundation, upon the 'user-illusion' of there being 'something it is like' to experience a unified reality that allows for conceptualizations and judgments to be brought in line with justification (see Chapter 2 for a detailed explication of this point). If Dennett denies this, and maintains that all experience is illusory, he faces the issue of explaining how P2's

judgments can be epistemologically justified without falling foul to the illusion of having a phenomenal experience in the process of this justification. Simply, Dennett cannot maintain his claim that our experience is ultimately a 'user-illusion' without first employing this same first-person 'user-illusion' of experience to justify his judgments, as, without this 'illusion', Dennett seems incapable of accounting for our capacity to accurately judge, and, in turn, our capacity to justify our judgments. Thus, our 'power to [accurately] judge' (McDowell 2008, p. 10) seems to be entwined to our phenomenal experiences in a manner Dennett (1991) fails to account for, and, in this oversight, Dennett's framework seems to risk becoming reliant upon both the very first-person reports that he deems to be unreliable, and the very phenomenal experiences that he is attempting to deny.

Further, even if we were to allow for P2's judgments to somehow avoid falling foul to the issues of first-person interpretive bias and/or reliance on phenomenal experience, I argue that Dennett's method would still fail to provide any semblance of a truly objective account of conscious phenomena by virtue of necessarily failing to consider *all* of the data contained within the beliefs and judgments of P1. In foregoing the first-person data in this way, I argue, in line with Nagel (1986), that true objectivity is lost, as upon failing to account for P1's judgment, we fail to achieve a complete account of reality (p. 118). Indeed, as Nagel (1986) articulates, in denying the efficacy of first-person perspective at the outset, the purportedly objective method not only seems to betray its own objectivity, but also necessarily leads us to the denial of the existence of phenomena that patently exist:

'A great deal is essentially connected with a particular point of view, or type of point of view, and the attempt to give a complete account of the world in objective terms detached from these perspectives inevitably leads to false reductions or to outright denial that certain patently real phenomena exist at all.' (Nagel 1986, p. 7)

In this respect, I argue that even if we were to forego the issues evident in maintaining the internal coherence of Dennett's method, Dennett still faces the issue of betraying his scientific 'neutrality' via positing the non-existence, or 'fictional' existence, of phenomenal experience at the outset (this is a contention shared by Carr 1998). As, in espousing the need for this 'free miracle', which dictates that as first-person reports cannot be reconciled with the scientific evidence, experience just *is* illusory; Dennett (1991) not only foregoes that which is most viscerally immediate, but, in so doing, seems to juxtapose the efficacy of our intuitions alongside the efficacy of the science that is only made possible by virtue of these intuitions. This is an issue for Dennett's framework, as the most fundamental human intuition does not entail a deep understanding of a physical, objective reality underpinned by physics, but instead entails a deep, foundational awareness of 'what it is like' to be oneself. It is an intuition that guides our earliest interactions and informs our later epistemic frameworks, and so in positing that because physical science cannot account for phenomenal experience we must accept that our deepest intuition just *is* an illusion, Dennett invites the question, which is more illusory: the physical, scientific, heterophenomenological method that denies phenomenal experience, or the phenomenal experience that makes the scientific method possible to begin with?

This question is underpinned by the notion, as outlined in chapter 2 of this thesis, that our justifications for inferences to the existence of an objective reality, and in turn Dennett's purportedly objective method, are tightly entwined with our phenomenal experiences in such a way that if judgments pertaining to these experiences are illusory, and experiences are themselves (even if only partially) the source of our epistemic justifications for the concept of both an objective physical reality and an objective scientific method, then Dennett (1991) seems to risk plunging his framework into the very illusionism he is attempting to explain. As an example, in writing this piece, my capacity to continue writing is predicated upon my working under, what Dennett (1991) might term, the 'user-illusion' of holding an experiential unity that contains, and grounds, a multiplex of unified judgments pertaining to my beliefs, desires, perceptions, and position in space-time. Yet, if Dennett (1991) maintains that the conscious experience that grounds these judgments is illusory, I contend that besides the fact that it seems obvious to ask 'Illusory to whom?' Dennett must also explain how the judgments that are grounded within the illusion of consciousness can be employed in the formation of his framework (as he himself suggests they are) without plunging the framework itself into an illusionism that undermines the efficacy of the scientific methods employed to substantiate his theory? As, if our judgments are implicitly contained within our apparently 'illusory' capacity for phenomenal experience, and Dennett (1991) employs judgments to underpin his framework, then every judgment we make becomes equally illusory, and Dennett's heterophenomenological method, contrary to its pursuit of a 'third person' neutrally objective science, becomes incapable of affirming the reality of anything.

Thus, I argue that Dennett cannot circumvent the issues faced by Sellars (1963), as in denying phenomenal experience, Dennett foregoes the experiential unity that underpins our 'power to judge' (McDowell 2008, p. 10), and, in so doing, foregoes the efficacy of the foundational, epistemic property that underpins his heterophenomenological method. So, it seems in his adamantly upheld conviction that conscious experience must be eliminated, Dennett unavoidably also eliminates the epistemic foundation of his framework. As a result, it seems Dennett's case for eliminativism collapses under the weight of the epistemic fragility that arises upon eliminating phenomenal experience, and so, in chapter 4, I explore the case for reductive representationalism, which, instead of eliminating experience, attempts to incorporate it into the physicalist framework.

Chapter 4

Reductive Representationalism

Reductive representationalism, or 'strong' intentionalism (for proponents see Dretske 1995, Tye 1995, Byrne 2002, Lycan 1996, Harman 1990) arose as an attempt to provide a physicalist framework capable of avoiding the problems inherent within eliminativism and mind-brain reductivism, via explaining how phenomenal content may be reduced to representational content, and in turn, reduced to the external properties of physical, intentional objects[20]. In contrast to 'weak' non-reductive intentionalism, of the kind upheld by Crane (2003) and Chalmers (2004), which remains firmly entrenched in the 'hard problem' via contending that phenomenal content is not reducible to represented content, the reductive representionalist maintains that the 'hard' problem is unequivocally solved upon reductively explaining the identity of phenomenal content within the representational content that accompanies the intentional interactions between the brain and its environment. An analysis of reductive representationalism shall act as the focus of this chapter. I shall begin with a brief explication of the theory, before concluding that the reductive representationalist's reliance upon qualia externalism produces two fundamental flaws, namely: the issue of explaining differences in phenomenal qualities, and the issue of accounting for our subjective, intuitive awareness in purely representational terms. Upon having explicated these issues, I shall conclude that the tendency of reductive representationalism to massage phenomenal character in order to fit the

[20] Intentional objects are here defined, in line with Martin (1998), as 'ordinary existing entities' (p.101), and I shall be using the term intentional objects interchangeably with 'physical objects' throughout. I also note here that Crane (2001) takes umbrage with this definition, contending that the 'intentional object is just the object (for some subject) of an intentional state or act' (p. 349). However, for my purposes in this chapter, it is not necessary to contest this point, as it seems clear that reductive representationalism (as a physicalist theory of the mind) is positing such intentional objects as 'ordinary existing entities' that objectively, physically exist independently of the mind.

reductive, intentional framework results in an untenable theory of consciousness that, ultimately, fails to overcome the problem it purports to solve.

Kriegel (2017) offers a widely applicable and cogent explication of reductive representationalism:

'There is a class of entities E and a type of relation R, such that (i) R is a non-phenomenal relation and (ii) for every phenomenally conscious state S, what makes S the phenomenally conscious state it is, and a phenomenally conscious state at all, is that S bears R to the members of E it does, and bears R to members of E at all.' (p. 4)

This formation is underpinned by the attempts of early reductive representationalists, such as Lycan (1996), to reduce the identity of phenomenal experience 'S' entirely to the brains 'relation to' or 'representation of' an intentional object 'E'. So that, if the representational content 'R' is identical to the phenomenally conscious state 'S', and 'R' is entirely reducible to 'E', the representationalist is able to reductively explain phenomenal experience 'S' by appeal to an objective, physical object 'E', and, in so doing, is purportedly able to provide a reductive account of phenomenal experience that avoids the issues of reducing phenomenal content to the brain and the problem of outright eliminativism faced by Dennett (1991). So, my experience of being sat atop this chair, and my perception of the black letters cascading on the screen as I feel the indent of my fingers on the keyboard, are all, under the reductive representational framework, experiences that are ultimately nothing beyond my brain's representation

of the content[21] contained within physically reductive entities 'E'. Thus, in this example, the keyboard, the chair and the letters, are all physical entities, which my brain, upon relating in the right way to these entities, forms a 'what it is like' sensation that represents the properties of these physical entities. The crucial move in the reductive representational account is to posit this 'what it is like' sensation as nothing beyond the content that occurs upon my brain accurately representing the content inherent within physical entities 'E'. Meaning that, the fundamental aim of reductive representationalism is to posit that 'phenomenal character is one and the same as representational content' (Tye 2000, p. 45), and, consequently, explain how phenomenal content is reducible to the external content of physical entities; such that any notion of there being 'something it is like' for me to experience the blackness of these letters on the screen is entirely explained in physically reductive terms by virtue of my brain's capacity to represent the actual black content of these physical entities. As such, the strength of this framework lies in its capacity to avoid having to explain how the identity of phenomenal content can be reduced to the physical brain, via attempting to reductively account for phenomenal qualities in external content.

This move to posit phenomenal experience as representational content, and represented content as reducible to the content that is implicit to physical entities is most championed in the works of Harman (1990), Tye (1995) and Dretske (1995), who employ a 'transparency thesis' as a means from which to strengthen their case for explaining phenomenal content as the brain's representation of the content inherent

[21] Here I am referencing content as the intentional content that represents the properties contained within a given intentional object. In line with my attempt to denounce reductivism, I am here referencing intentional objects as physically reducible entities, in line with Martin (1998). However, as Crane (2001) explains, such intentional objects need not necessarily be physical or even a property at all.

within physical properties. Harman (1999) typifies this transparency thesis neatly in his contention that:

*'When you see a tree, you do not experience any features as intrinsic features of your experience. Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict you will find that the only features there to turn your attention to will be features of the presented tree.' (*p. 251)

Thus, according to advocates of this thesis, reductive representationalism is affirmed by virtue of my inability to introspectively focus my attention upon an intrinsic feature of my experience of the blackness of these words, for example, without first focusing upon the content of 'blackness' contained within this intentional object. Meaning that, any attempt to disentwine my phenomenal experience of 'blackness' from the physical object itself is impossible, as in separating myself from the intentional object, I am no longer truly aware of the experience I am attempting to draw my attention to. As such, according to advocates of reductive representationalism and the transparency thesis, the black quale accompanying my experience of these letters is nothing beyond what my brain represents the content of these letters as actually being, meaning that when the representation of the content of a physical entity is veridical, my phenomenal experience of 'blackness' is entirely accounted for by the mind-independent, physically reducible 'colour-properties' of this intentional object. This argument results in the representationalist contending that we 'see right through it (experience) and onto the objects and properties in the external world' (Batty 2010, p. 103), as 'in turning one's mind inward to attend to the

experience, one seems to end up concentrating on what is outside again, on external features or properties' (Tye 1995, p. 30).

This has led certain reductive representationalists, such as Dretske (1995), to champion qualia externalism, and deny the import of Nagel's (1974) notion of equating phenomenal experience to an intractable, wholly subjective 'what it is like' sensation. Instead contending that Nagel's claims involving the purported impossibility for me to know 'what it is like' to be a bat are fallacious, as in order to know 'what it is like', we must look no further than the content of the properties inherent within the bats external environment. Dretske (1995) doesn't directly reference Nagel's bat, but instead employs the notion of a parasite, contending:

'If you know what it is to be 18 °c, you know how the host feels to the parasite. You know what the parasite's experience is like as it 'senses' the host. If knowing what it is like to be such a parasite is knowing how things seem to it, how it represents the objects it perceives, you do not have to be a parasite to know what it is like to be one. All you have to know is what temperature is…To know what it is like for this parasite, one looks, not in the parasite, but as what the parasite is 'looking' at'- the host.'
(p. 83)

Thus, Dretske (1995), in line with Tye (1995) and Lycan (1996), attempts to disentwine phenomenal content from the physical brain, and instead posit the intentional objects inherent within an environment as entirely accountable for the brain's representational, and in turn, phenomenal content. Contending that, if our phenomenal experience of heat is 18 °c, we should not look within the brain to

distinguish why this is so, but instead should look to the environment in which the brain represents this content. Thus, the reductive representationalist removes the need to internally explain phenomenal 'what it is like' sensations, and instead posits that: as phenomenal content is 'one and the same as representational content' (Tye 2000, p. 45), such that to be the subject of 'a state with a certain felt or phenomenal quality is to be the subject of a state that represents a certain *external* quality' (Tye 2000, p. 162), we should posit the content of our phenomenal 'what it is like' sensations as reducible to the external properties of physical objects. So, on this externalist view, the phenomenal character of an experience is not contained within the neurological makeup of a given brain state, but is instead indicative of the 'character of the reflectance-type (of the physical properties) which that brain-state tracks' (Pautz 2003, p. 7). Employing this reduction of phenomenal content to the content inherent within physical properties allows the representationalist to avoid the 'hard' problem plaguing qualia internalism, whilst reductively explaining phenomenal experience via predicating their framework upon the contention that 'qualia ain't in the head' (Tye & Byrne 2006, p. 241) [22].

Thus, according to reductive representationalism, the phenomenal experience of person 1's (P1's) experience is identical to P2's experience *iff* they are both representing the same perceptible properties of object 'X'. Meaning that any difference in phenomenal experience must be wholly accounted for by a

[22] Chalmers (2004) posits that this move to externalize qualia is necessary for the reductive representationalist, as attempting to provide a reductive representational framework not predicated upon externalism would result in, what Chalmers (2004) terms, a 'inconsistent triad' (p. 166) of simultaneously upholding (i) reductive representationalism (ii) internalism about phenomenal qualities and (iii) externalism about content. This means that the representationalist can either advocate reductivism alongside externalism or advocate non-reductivism alongside internalism; he or she cannot be a representational reductivist whilst maintaining internalism about phenomenal content, without risking either advocating a thesis that is not predicated upon representationalism, or no longer advancing a reductive theory of the mind.

representational difference that is reducible to a difference in the properties of a given object. This, however, seems troublesome, as any move to externalise qualia is predicated upon the reductive representationalist making an ontological claim about what phenomenal experience actually is, namely: 'one and the same as representational content', which is, at foundation, 'one and the same' as the content contained within the properties of intentional objects (Tye 2000, p. 45). In making this ontological claim about the externally reducible nature of phenomenal experience, I argue that reductive representationalism reveals a platform from which to espouse two fundamental arguments against the qualia externalism that underpins its reductivism, namely: the problem of explaining differences in phenomenal character arising from the same intentional object, and the problem of accounting for the unified, intuitively aware locus of experience that seems implicit to phenomenal experience. I contend that, if these arguments hold, I provide convincing reasons to doubt the externalism underpinning representational reductivism and, in so doing, strengthen the case for the irreducibility of phenomenal information.

The first of these arguments is established by Pautz (2006) and Cohen (2009), who attempt to posit, what I term, an 'inverse argument for multiple realizability', via highlighting the difficulty for qualia externalists to explain how different phenomenal experiences may arise from a singular intentional object, that, according to the qualia externalist, should produce no such disparity in experience. As such, whilst reductive representationalism purportedly avoids the standard problem of 'multiple realizability'[23] (See Putnam 1967) via explaining how disparate physical organisms may produce the same phenomenal content in their veridical representations of the

[23] Simply, this is the problem of accounting for how distinct physical identities may produce the same phenomenal content, and is traditionally espoused as an argument against reductivism.

external properties of intentional objects, Pautz (2006) and Cohen (2009) attempt to topple the purported strength of the representationalist's capacity to avoid the problem of multiple realizability via inverting the problem entirely. Thus, whilst the standard problem is concerned with the problem of distinct physical brains experiencing the same phenomenal content, Pautz's formation is concerned with the problem of distinct phenomenal content arising from the same physical object.

As such, Pautz (2006) is contending that if the representational externalist maintains their claim that physical properties implicitly contain some informational content that wholly constitutes our representational, and in turn, phenomenal content, then we should be entirely unable to conceive of a scenario in which distinct phenomenal experiences arise from veridical representations of the same physical properties. Meaning that:

'No matter what world you go to, the unitary red brain state tracks the unitary red reflectance in the actual world. So, no matter what world you are in, if you are in the unitary red brain state, you bear the Rigidified Tracking Relation to the unitary red reflectance, and so, given physicalism about Q-properties (phenomenal properties), the colour unitary red.' (Pautz 2006, p. 218)

However, as Pautz (2006) highlights in his 'Maxwell and Twin-Maxwell' thought experiment, we *can* conceive of disparate phenomenal experiences arising from the same physical properties. This example involves Maxwell (X) as a standard, human perceiver in this world, and 'Twin-Maxwell' (Y) as an occupant of a twin-world, in which colour perception has evolved to produce no less optimal, but slightly different,

'post-receptoral wiring' (p. 213) to that of X. The argument suggests that, whilst both X and Y share identical retinal configurations, their different post-receptoral wiring produced disparities in the way in which photoreceptor signals are processed, which conceivably results in a representational, and in turn phenomenal, difference in how X and Y experience the same intentional object. If this argument is upheld, it presents an issue for qualia externalism, which dictates that it is impossible for distinct experiences to arise from veridical representations of the same intentional object.

However, in order to understand this, it seems prudent to first briefly outline the scientifically accepted explanation for colour vision, so as to elucidate a deeper understanding of the consequences of this thought experiment. Simply, upon viewing a given intentional object, light energy enters our retina, and is then mapped to a series of photoreceptor cells that convert this light energy into electrical signals, which are then transmitted 'to a series of post-receptoral elements which process these signals and send them to second order retinal neurons' (Kremers et al 2016, p. 46). So, more simply, light energy is converted into an electrical signal, which is then transferred to the brain via a series of nerve fibres. At this stage, the brain's post-receptoral processing maps the electrical signal of the nerve fibres to a given photoreceptor, and this result is then 'coded in the train of neural impulses sent by retinal ganglion cell axons to the higher visual centres of the human visual system' (Kremers et al 2016, p. 46) which account for our representational, and in turn, phenomenal experiences of colour. This process is eloquently conveyed as follows:

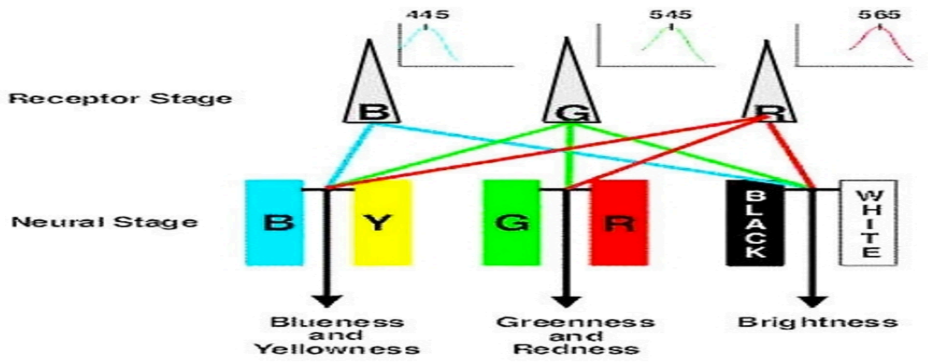


Figure 1 *A Model for Human colour vision* (Source: Kalloniatis & Luu 2007).

So, as the diagram conveys, the 'receptor stage' consists of two processes, the first involves signals being transmitted from photoreceptors to the brain's post-receptoral 'cones', and the second involves these post-receptors matching a signal to a given neurological output (or channel) so as to produce our experience of colour. Now, imagine that our evolutionary cycle had produced post-receptoral wiring that, upon processing the electrical signal of '445' (as seen in the diagram above), only mapped this signal to one neurological channel, instead of the two depicted in the diagram, so that instead of experiencing blue as a particular hue that produces our sensation of 'light' blue, we only experience signal '445' as a blueness devoid of a particular brightness or hue. If this is conceivable, it is conceivable that this difference in post-receptoral wiring has caused a difference in our phenomenal experience of blue, and therefore it is equally conceivable that phenomenal content is not contained within external objects, but is instead a property of how the brain processes electrical signals. With this established, we are in a position to return to the thought experiment championed by Cohen (2009) and Pautz (2006).

The key contention is that, if Maxwell and Twin-Maxwell hold distinct post-receptoral wiring whilst also being evolutionarily optimised, then the functional

differences inherent in the way in which they process light will produce phenomenal differences in the way they experience colour. With this as a grounding, Pautz's (2006) argument highlights that, according to representational externalism, both X and Y should experience identical phenomenal content upon representing the same physical properties of a given object 'E', however, due to variations within their 'post-receptoral' neurological makeup, it is conceivable that X represents E as instantiating a distinct hue that causes a phenomenal experience of 'orange', while Y represents E as instantiating the colour content of 'redness'. So that the differences in post-receptoral wiring creates differences in how signals are mapped to neurological channels, which, in turn, creates differences in phenomenal experience. Allen (2016) describes this as follows:

'In Maxwell's case a particular retinal response activates just one opponent (neural) channel, whereas in Twin Maxwell's case it activates two opponent channels. So Maxwell and Twin Maxwell have phenomenally distinct experiences: an object that Maxwell sees as instantiating a unique hue, Twin Maxwell sees as instantiating a binary hue.' (p. 75)

Thus, the representations, and, in turn, phenomenal content gleaned from the perceptible properties of object 'E' are varied in a manner that seems to contradict the notions of representational externalists, which leads Pautz (2010) to contend that 'experiential properties are very well correlated with neural properties and very poorly correlated with external properties' (p. 34). Indeed, both Cohen (2009) and Pautz (2006/10) take this argument as evidence for the alleged falsity of qualia externalism, as our representations and, in turn, phenomenal experiences, seem

dependent upon what happens in the brain, not the properties of our external environment. This seems intuitive, as qualia externalism's commitment to externalising phenomenal content leaves us unable to account for the conceivable contention that, in the case of Maxwell & Twin-Maxwell, there *is* a disparity in phenomenal experience upon representing the same intentional object.

In an attempt to defend qualia externalism, Tye & Byrne (2006) suggest that any disparity in represented content is explained by a disparity in the extent to which these representations are veridical. Thus, Tye & Byrne deny that there is any need to explain cases such as Maxwell/Twin-Maxwell, as Maxwell's (X) representation of the intentional object (E) as withholding orange content is simply a more veridical representation of the actual properties of E than that offered by Twin-Maxwell (Y), meaning that, ultimately, Y's phenomenal experiences are illusory in a way X's are not. As such, according to Tye & Byrne (2006), phenomenal content *is* accounted for externally, and any perceived disparity in experience is simply a matter of the brain incorrectly representing this external content. So, the crucial move here is an attempt to firmly deny the brain's capacity to internally produce phenomenal content, and maintain that any case in which the brain *seems* to produce this content is a case in which the brain is producing nothing beyond a warped, illusory depiction of content that *is* reducible to external, intentional objects. Further, this focus upon the illusory nature of Y's experiences leads Tye & Byrne to question how Pautz (2006) and Cohen's (2009) contentions can be reconciled alongside evolutionary theory, as it seems difficult to explain 'how Twin Maxwell could have evolved so as to systematically misrepresent the colours of things' (Tye & Byrne 2006, p. 253).

All such objections, however, seem predicated upon the assumption that Twin-Maxwell (Y) *is* necessarily misrepresenting the perceptible properties of the object (E); which seems to demand that the evolutionary cycle for Maxwell (X) has somehow produced a representational, and evolutionary, optimisation that is not reflected in his twin (making X's representations optimal to survival in a way that Y's are 'sub-optimal'). These claims seem to either misrepresent Pautz's arguments or misrepresent the Darwinian theory of natural selection, as the thought experiment demands that both X and Y epitomise systems that have optimised their capacity for survival, which means that, over the course of natural selection, this survival optimisation must necessarily have been predicated upon both X and Y consistently *not* misrepresenting their environments.

Hence, far from dismissing Twin-Maxwell's representations as 'sub-optimal' or non-veridical, we must accept that, if both X and Y are evolutionarily optimised, neither X nor Y are truly 'misrepresenting' the 'colours of things', because the representations of X and Y must be equally veridical in terms of survival optimisation, as Pautz (2006) contends: 'the situations of Maxwell and Twin Maxwell are perfectly symmetrical…[so] that optimal [survival] conditions obtain in Maxwell's situation and Twin Maxwell's situations' (p. 223). If this is upheld, it seems only logical to question why exactly the representational externalist believes that Maxwell's representations are any more veridical than those of Twin-Maxwell, because, if both are evolutionarily optimised, then our post-receptoral wiring renders any attempt to truly *know* a veridical representation as futile. Meaning that, if it is at least possible for two distinct individuals to represent, and in turn, experience the same object differently, whilst maintaining identical evolutionary optimisation, then it seems

logical to conclude that phenomenal experience is not reducible to external properties in a given environment. For, if we can at least conceive of two evolutionarily optimised individuals holding distinct phenomenal experiences whilst representing the same external content, then we are presented with a good reason to deny qualia externalism. By means of response, Tye & Byrne (2006) seem to fall back on the notion that 'there is no obvious reason to suppose [that the evolutionary optimisation of Twin-Maxwell] is metaphysically possible', as it would be metaphysically impossible for a 'product of natural selection…operating under the same laws as Maxwell with a similar kind of visual system' to produce different representational and, in turn, phenomenal content in relation to an object 'E' (p. 253). In order to substantiate this claim, it seem that Tye & Byrne (2006) must maintain that Twin-Maxwell is wholly inconceivable. However, this in itself seems difficult to establish, as it not difficult to imagine that our twins on another possible world may have reached a state of evolutionary optimisation that has provided survival strategies that are equivalent to our own, whilst, for example, experiencing the dark blue hue of electrical signal '445' as a very slightly lighter blue hue than the experience the human species has of signal '445'. This minimal conception of a system's evolutionary optimisation (in some possible world) remaining intact upon experiencing the hue of a singular electrical signal very slightly differently to those on earth is all that is required in order to conceive of Twin-Maxwell, and, as such, I posit that there is no coherent reason for Tye & Byrne to maintain that Twin-Maxwell is not conceivable. Indeed, if they do persist in this claim, they seem to commit themselves to the modal fallacy of leaping from possibility to necessity, by upholding the presumptuous inference that the prerequisites for achieving evolutionary optimisation in this possible world must necessarily occur in order to achieve

evolutionary optimisation in all possible worlds. With this in mind, I maintain that it is important to reiterate Chalmers' (2010) distinction between an ideal rational reflection (or primary conception) that would, in this case, pick out any referent with evolutionarily optimized qualities, and a prima facie belief (secondary conception) that would assert that a referent's 'evolutionary optimization' necessarily depends upon experiencing signal '445' exactly the same as those on earth. In this case, I argue that Tye & Byrne (2006) make the mistake of overemphasising the a posteriori, prima facie belief that evolutionary optimisation can only occur upon experiencing the signal 445 as a human does, whilst understating our potential to employ ideal, primary conceivability to access a possible world in which the concept of evolutionary optimization simply picks out any referent with the quality of having optimized survival strategies. Upon employing this ideal, primary, conceivability, I argue that Twin-Maxwell is conceivable, and, in line with Chalmers (2010), I maintain that, upon conceiving of this, we access a 'primarily possible' world in which Twin-Maxwell is verified by some centred being, and, in so doing, access a 'secondarily possible' world in which the existence of Twin-Maxwell is satisfied. As such, I employ the following adaption of Chalmers' (2010, p. 144) conceivability argument (as has been employed and explicated in detail within chapter 1 of this thesis) to account for the leap from the (primary) conceivability of Twin-Maxwell to the metaphysical possibility of a possible world in which the existence of Twin-Maxwell is satisfied[24].

1. It is conceivable that Twin-Maxwell achieves evolutionary optimisation whilst experiencing the colour content of E slightly differently to Maxwell.

[24] For a more detailed explication of this particular conceivability argument, see my formation of Chalmers' (2010) reformulated conceivability argument in chapter 1.

2. If this is conceivable, the evolutionary optimisation of Twin-Maxwell is primarily possible.
3. If the evolutionary optimisation of Twin-Maxwell is 1-possible, then it is secondarily possible.
4. Therefore, the evolutionary optimisation of Twin-Maxwell is metaphysically possible in some possible world.

The first and most damning consequence of this argument is that Tye & Byrne (2006) are left incapable of truly explaining phenomenal content. As, if it is conceivable that Twin-Maxwell holds a distinct experience alongside a veridical representation of an external property, then it seems equally conceivable that phenomenal content contains a degree of subjective, experiential awareness that is ontologically irreducible to objective, external properties. Meaning that, ultimately, in failing to address why Twin-Maxwell is not conceivable, Tye & Byrne (2006) also fail to explain phenomenal content in its entirety (this argument shall be returned to shortly). Further, if it is conceivable, and therefore metaphysically possible, that both X and Y are evolutionarily optimised whilst their neurological makeup produces disparate experiences of E, it is relatively easy for Pautz (2006) to defend both his claim that internalism is a more cogent theory than externalism (although physicalist internalism still seems to inevitably fall foul to the hard problem), and his claim that: if representational externalism demands that phenomenal content is externally reducible to the properties of a given intentional object, so that X and Y necessarily produce identical phenomenal experiences in their veridical representations of object 'E', then the Maxwell/Twin-Maxwell thought experiment goes some way in proving that representational externalism is false, and so too, goes some way in pacifying the

attempts of those reductive representationalist who wish to reduce phenomenal content to physical content.

Indeed, if Pautz's argument is upheld, the qualia externalism underpinning representational reductivism is shown to be false, and so the entire framework of reductive representationalism, of the kind upheld by Tye (1995), Byrne (2002), Dretske (1995) and Lycan (1996), collapses. At this stage, reductive representationalism finds no relief in somehow aligning representationalism to reductive internalism (this is the solution Pautz 2006 seems to champion), because externalism arose precisely out of the difficulty of internally accounting for phenomenal content in terms of the physical brain; in fact, this is the very foundation of Chalmers's (1996) hard problem (see chapter 1.1/1.2 for an explication of this point). So, the underlying motive for externalising qualia, and employing the brain as a machine capable of representing this external content, is one of avoiding the difficulty of internally explaining qualia by reductively identifying it with the brain, and, as a result, reductive internalism is denied even more swiftly than qualia externalism. As such, this leaves the reductive representationalist with very few options other than to deny the metaphysical possibility of Pautz's argument (see Tye & Bryne 2006). However, as we have explored, this in itself seems insufficient.

With this established, I move on to address, what I deem to be, the most fundamental issue underpinning the representational externalist framework, namely: the problem of accounting for the phenomenal experiences that seem to be beyond encapsulation within external properties, and, in turn, seem wholly non-representational. The contemporary debates in this area focus upon the non-representational nature of

moods and pains, and, in the remainder of this chapter, I shall posit that these non-representational phenomena imply a unified intuitive, subjective awareness of 'what it is like', which is entirely distinct from the objective, sensory phenomenal information that seems to be the focus of representational externalism. Meaning that, even if we were to forego Pautz's (2006/2010) convincing dismissal of qualia externalism, the reductive representationalist would still fail to truly capture the fundamental nature of phenomenal character.

Thus, I contend that the strongest reason for denying the adequacy of representational externalism is found upon positing qualitative content as containing something beyond the purely sensory experience that accompanies our representations of intentional objects. Here, I am referencing the sort of subjective content that seems to be most evident upon being experientially aware of a tranquil pleasant mood, or a particularly dull and persistently unpleasant pain, as it is experiences of these kinds that seem entirely irreducible to representational content. Indeed, it is these seemingly non-representational experiences of qualitative content that present a deep threat to the reductive representationalist framework, resulting in us asking questions such as: if all phenomenal content is accounted for externally, what is my ubiquitous awareness of debilitating anxiety representative of? Such questions seem to reveal the fundamental issue with reductive representationalism, as whilst Tye (1995) and Dretske (1995) may argue that the physical black properties of these letters account for my phenomenal experience of 'blackness', it seems clear that these representations of external properties can in no way account for my being phenomenally aware. I argue that the more standard responses to this issue, such as, for example, that physical pains may represent damage to body tissues; or that anxiety is more akin to a

propositional attitude; or perhaps that anxiety is representative of chaos, all fail to provide befitting explanations for how represented, informational content makes the transition to content that we are subjectively, and internally, experientially aware of. Indeed, if the representationalist response to this particular problem is to simply fall back on the notion that the phenomenal quality experienced during anxiety is itself reducible to external, informational content, then it seems that the phenomenal quality involved in the subjective awareness of undergoing anxiety is lost, or 'morphed into something else' (Smith 2011, p. 361) that is itself an explanation for something other than the subjective phenomenal character that the representational account is attempting to explain. So, we are left with the troublesome problem of explaining how the chaotic representations, which perhaps may account for anxiety, are themselves capable of bridging the divide between objective, represented content and subjective states in which we are aware of undergoing this representational content. And so, we need an explanation for how representational content explains its experiential capacity without first providing an account of the experiencer for whom this represented content becomes experiential. Indeed, in this respect, I argue that upon positing an explanation for phenomenal experience, we must provide an explanation for both how experiential content arose, *and* an explanation for how we exist as experientially aware experiencers of this represented content[26]. As Strawson (1994) attests: 'A subject of experience…is something that must exist whenever there is experience, because experience is necessarily experience-for' (p. 133). This distinction seems of particular importance, because we know that conscious experience exists not because of the qualities of representations, but because we are

[26] The standard anti-physicalist responses avoid this issue by positing that a phenomenal property is experiential in such a way that whenever we posit an experience, we simultaneously posit an experiencer for whom the attribute of awareness may be posited (See Strawson 1994 for a deeper articulation of this point).

directly aware of our experiences, and, as a result, a theory of conscious experience must account for the subjective, experiential awareness that has allowed us to articulate the hard problem.

Hence, at foundation, the representational externalist faces the problem of accounting for how a purely physical, reductive representational framework, predicated upon externalising phenomenal content, may account for our subjective, experiential awareness of 'what it is like' to undergo a given representation as a unified experiencer. This distinction between phenomenal experience, construed as an experientially unified awareness of 'what it is like', and phenomenal content, construed as an objective, sensory, representational quality, is typified eloquently by Gibson (1979):

'*Direct perception is what one gets from seeing Niagara Falls, say, as distinguished from seeing a picture of it. The latter kind of perception is mediated. So when I assert that perception of the environment is direct, I mean that it is not mediated by retinal pictures, neural pictures, or mental pictures. Direct perception is the activity of getting information from the ambient array of light. I call this a process of information pickup that involves the exploratory activity of looking around, getting around, and looking at things. This is quite different from the supposed activity of getting information from the inputs of the optic nerves, whatever they may prove to be.'* (p. 139)

For the most part, the reductive representationalist fails to recognise the importance of this distinction between mediated, objective, physically reducible sensory information

and unmediated, subjective states of awareness. Often simply contending that 'producers and consumers' of representations are required in order to organise and causally unify potentially disparate representational content (Millikan 1990, p. 156), but offering very little explication for how the unified locus of experiential awareness, which seems implicit to this 'consumer' of representations, can be accounted for within the framework of reductive representationalism. Indeed, as I explicated in chapter 2 of this thesis, a capacity for unification of this kind seems to only be achieved upon positing a phenomenally unified, awareness of 'what it is like' to undergo these potentially disparate representations. This, of course, leaves us asking: how can a purported physically reductive solution to the 'hard' problem of consciousness be upheld without accommodating for the phenomenal awareness that seems so fundamentally implicit to a common-sense depiction of what consciousness actually is.

Tye (1995) seems to recognise this problem, and attempts to employ his notion that phenomenal states are Poised, Abstract, Non-conceptual, Intentional Content (PANIC) in an attempt to account for the transition from unconscious, represented content to a subjective state of conscious awareness. Whilst this move emboldens his representational externalism by qualifying the abstract, non-conceptual nature of his intentional content (thus avoiding issues of hallucinations and the issues faced by Sellars 1963, see chapter 2), the true importance of PANIC lies in Tye's (1995) notion of 'poise'. As, it is this appeal to 'poised' content that enables Tye to unite his externalism with a functionalism, which, he argues, allows him to bridge the gap between a representational mental state and a phenomenally conscious mental state. Tye (2000) describes the condition of being 'poised' as follows:

'This condition is essentially a functional role one. The key idea is that experiences and feelings, qua bearers of phenomenal character, play a certain distinctive functional role. They arise at the interface of the nonconceptual and conceptual domains, and they stand ready and available to make direct impact on beliefs and/or desires. For example, how things phenomenally look typically causes certain cognitive responses—in particular, beliefs as to how they are if attention is properly focused. Feeling hungry likewise has an immediate cognitive effect, namely the desire to eat. In the case of feeling pain, the typical cognitive effect is the desire to protect the body, to move away from what is perceived to be producing pain. And so on. States with nonconceptual content that are not so poised lack phenomenal character.' (p. 62)

So, the contention is that representational content becomes conscious upon being integrated into a functional cognitive system that is 'poised' to produce beliefs and desires as a result of this represented content. As Byrne (2002) puts it: 'a pang of hunger, say, is poised just in case it stands "ready and available to have a direct impact" on some beliefs and/or desires' (p. 11). The problem with this, however, is that whilst Tye (1995) attempts to strengthen the case for representational externalism and functionalism by combining the virtues of both, his framework must ultimately still confront the problem of explaining how this 'functional-representationalism' may account for the transition from objective, external properties to our ubiquitous, subjective awareness of 'what it is like' to *be* this given subject (this problem is also noted by Wheeler 2010, p. 266 and Chalmers 2004, p. 163). Such an explanation, I argue, shall forever elude Tye (1995), as his reductive representationalism demands

that phenomenal content be entirely contained in the properties of external objects, which means that Tye is entirely reliant upon his 'poised' functional state in order to convert this external, phenomenal content into internal, phenomenal awareness. A transition of this kind unequivocally overstretches functionalism, however, as our ubiquitous awareness of 'what it is like' encompasses something beyond that which is accountable within purely functional terms (this is the 'zombie argument' espoused by Chalmers 1996, see chapter 1.1), and so, at best, all Tye (1995) can maintain is that 'poise' somehow exists as a prerequisite for phenomenal *content*, but in no way acts as an explanation for phenomenal *awareness*.

Thus, the disparity between representational content and phenomenal awareness seems to persist, and Tye is left unable to explain how this external content may produce the internal states of 'being aware of a general sense of buoyancy, of quickened reactions, of somehow being more alive' (Tye 2002, p. 144) that Tye (2002) himself acknowledges as 'experienced qualities of oneself' (p. 144). Simply, Tye's framework is left incapable of establishing a convincing argument for why functionalism may be used to explain how external content produces states of internal awareness, as it seems in order to explain these states Tye must posit an internal, experiential awareness of 'what it is like' that seems entirely incompatible with functionalism and reductive representationalism, or indeed any physicalist account.

As a result, Tye's thesis offers no distinct explanatory value to the problem of conscious experience, by virtue of ultimately failing to overcome the issue of explaining phenomenal awareness in purely representational terms. Further, this failure to capture the fundamental nature of phenomenal experience seems to neatly

typify the issues underpinning all reductive representationalist accounts, as all such frameworks seem predicated upon an attempt to massage phenomenal character in order to fit the reductive, physicalist framework. As I have attempted to explicate within this chapter, these attempts to manipulate phenomenal experience in order to explain its reduction to physical, external properties result in highly counter-intuitive conclusions that seem to either contravene our understanding of evolutionary theory, or fail to capture the locus of experiential awareness that seems so implicit to phenomenal experiences. As such, contrary to its intention of providing a physicalist account of phenomenal content, the explanatory failures of reductive representationalism leave us looking to anti-physicalism as a potential remedy to the issues of misrepresentation that arise upon attempting to place phenomenal awareness within the framework of physical reductivism.

Chapter 5

The Phenomenal Concept Strategy

As my account of the physicalist solutions to the problem of consciousness draws to a close, we witness a common pattern emerging amidst the failed attempts to reductively account for phenomenal experience. All of which seem to acknowledge the difficulty of fully accounting for experiential qualities, and so attempt to massage our conception of what phenomenal experience actually is in order to produce a framework in which this warped depiction of quasi-consciousness is purportedly physically explained. As I have explicated throughout, all such theories ultimately fail precisely because any misrepresentation of phenomenal experience leaves the ontological gap underpinning the hard problem wholly intact, which, as I explained in the previous chapter, leaves us looking to anti-physical metaphysical frameworks capable of fully accounting for the identity of experiential qualities.

In response to these failings, contemporary proponents of physicalism have attempted to save their framework by positing that this apparent ontological gap underpinning physical and phenomenal content can be avoided upon establishing that the 'hard' problem is nothing beyond an epistemic disparity between physical and phenomenal *concepts.* Thus, proponents of the 'phenomenal concept strategy' (see Loar 1990/9 Carruthers & Veillet 2007, Balog 2009/12, Papineau 2002) contend that if the conceptual, epistemic problems underpinning the hard problem can be solved in a manner compatible with physicalism, then 'sufficient doubt' has been cast unto the 'the anti-physicalist arguments that link semantic/epistemic gaps with ontological gaps' (Balog 2012, p. 9). An analysis of this attempt to avoid the metaphysical issue

of accounting for the ontology of non-spatial experiential phenomena, and in turn, uphold an appeal to monistic physicalism via reducing the problem of consciousness to a problem within our epistemic/conceptual frameworks, shall act as the focus of this chapter. I begin with a brief explication of the recognitional and constitutional formations of the strategy, before concluding that, if, as explicated in chapter 2, the 'easy' problem of epistemic justification cannot be disentwined from the 'hard' problem, then any conceptual strategy must ultimately fall back on a metaphysical dualism that leaves the ontological gap wholly intact (similar arguments are employed by various philosophers of mind, see Livingston 2013, Tartaglia 2015, Furst 2008). By the end of this chapter, I hope to have compounded my case for dismissing physicalistic monism, and re-affirmed the need for an anti-physicalist solution to the hard problem that is capable of accommodating the troublesome nature of the ontological gap.

As my starting point, I employ Chalmers's (2010) elucidation of the phenomenal concept strategy, or, what he terms, 'type B materialism':

'Proponents put forward a thesis C (the phenomenal concept strategy) attributing certain psychological features — call these the key features — to human beings. They argue (i) that C is true, i.e. that humans actually have the key features; (ii) that C explains our epistemic situation with regard to consciousness, i.e. that C explains why we are confronted with the relevant distinctive epistemic gaps; and (iii) that C itself can be explained in physical terms, i.e. that one can (at least in principle) give a materialistically acceptable explanation of how it is that humans have the key feature.' (p. 311)

Thus, the arguments underpinning attempts to employ phenomenal concepts in order to strengthen the case for physicalism seem best elucidated as follows:

If we experience an epistemic gap between physical and phenomenal concepts, *and*

1. All cognitive states are physical states.
2. Phenomenal and physical concepts are states of cognition.
3. Thus, both physical and phenomenal concepts must, as states of cognition, ultimately reference physical states.

Then, there must be an explanation, which is consistent with physicalism, for *why* we experience a conceptual and epistemic gap.

So, as distinct from type A materialism, which appeals to an illusionism of the kind explicated within chapter 3 of this thesis, type B materialism attempts to 'locate the gap in the relationship between our *concepts* of physical processes and our *concepts* of phenomenal processes' (Chalmers 2010, p. 305) so as to explain why many philosophers ostensibly *mistakenly* conclude that the explanatory gap at the core of the hard problem is an ontological gap. Thus, the central contention underpinning the phenomenal concept strategy (henceforth referenced as the 'PCS') is that the purported 'hardness' of the hard problem is accounted for entirely by the dualistic nature of our concepts, not the dualistic nature of reality. Meaning that our tendency to conclude that our phenomenal experience of 'what it is like' to experience pain is ontologically distinct from pain identified as physical 'C-fibre stimulations', or our

capacity to conceive of our physical duplicate being devoid of phenomenal experience, arise solely because our phenomenal concepts are epistemologically isolated from our physical conceptualizations. Thus, the PCS, as a physicalist strategy, appeals to conceptual dualism (phenomenal concepts exhibit certain features that epistemologically isolate them from physical concepts) whilst firmly maintaining ontological monism.

This framework is underpinned by an appeal to disentwine our *concepts* of phenomenal properties from the *properties* themselves. So that, our concept of phenomenal experience (Q) is nothing beyond the way in which we conceptualize and think about Q, which is distinct from the properties of Q that exist isolated from said thoughts and conceptualizations. Thus, whilst the concept of Q I have employed throughout this thesis is one of Q existing as a unified, intuitive awareness of 'what it is like'; proponents of the PCS (see Loar 1990, Carruthers & Veillet 2007) highlight that, ultimately, all this conveys is that I hold certain physically reducible cognitive tools that allow me to infer concepts pertaining to the properties of Q, not that these concepts in any way reveal the fundamental nature of the properties of Q. Thus, the PCS employs an appeal to the epistemic primacy of our concepts, as, upon employing phenomenal concepts, our cognitive tools purportedly inhibit our capacity to truly know the properties of Q as a fundamentally physical structure devoid of the conceptualizations we employ. If this is upheld, advocates of the PCS conclude that the explanatory gap underpinning the hard problem is accounted for by disparities within our physical and phenomenal *concepts*, not necessarily by disparities within physical and phenomenal *properties,* so that the problem is immediately reduced to an issue of accounting for this conceptual, not ontological, dualism. Hence, Loar (1990-

9) and Carruthers & Veillet (2007) contend that this enables us to uphold an appeal to monistic physicalism, as the apparent ontological gap underpinning the hard problem is immediately reduced to an epistemic disparity within our conceptualizations that is entirely consistent with physicalism. As Loar (1990) attests: disentwining our concepts of Q from the properties of Q that exist independently of our conceptualizations leaves the anti-physicalist only able to affirm an epistemic disparity in our conceptualizations, and provides no basis from which to argue for an ontological distinction between phenomenal and physical properties (p. 203).

It is this move to supplant the ontological problem with an epistemological problem that is the strength of the PCS as a physicalist framework. As, upon reducing the problem to a disparity in our concepts, the physicalist avoids having to account for the ontology of non-physical experiential properties, and instead is able to uphold an appeal to monistic physicalism, via reducing the hard problem to the problem of epistemic isolation that occurs upon attempting to plot our concepts of phenomenal properties alongside our concepts of physical properties. This appeal to the epistemic isolation of phenomenal concepts seems implicit to all attempts to bring physicalism in line with PCS, with all contending that phenomenal concepts necessarily contain certain 'special features' (Tye 1999, p. 707) that are both largely overlooked by the anti-physicalist movement and wholly account for the explanatory gap that underlies the hard problem. Further, according to Loar (1990), once our understanding of the special epistemic isolation of phenomenal concepts is established, we will glean a means from which to 'take the phenomenological intuition at face value, accept introspective (phenomenal) concepts and their conceptual irreducibility, and at the same time take phenomenal qualities to be identical with physical-functional

properties of the sort envisaged by contemporary brain science' (p. 196). Thus, at foundation, the PCS rests upon an attempt to cast 'sufficient doubt' (Balog 2012, p. 9) unto the anti-physicalist appeal to account for the explanatory gap in ontological terms, via upholding the central claim that *iff* the epistemic isolation underlying the disparity between phenomenal and physical concepts can be accounted for in a manner compatible with physicalism, then the hard problem is solved without having appealed to an ontological gap.

As such, the efficacy of the PCS as a physicalist strategy seems to rest entirely upon its capacity to explain *why* we witness the 'special features' (Tye 1999, p. 707) that account for the epistemic isolation of phenomenal concepts, and, more importantly, *how* the epistemic disparity between phenomenal and physical concepts arose if, at foundation, all concepts reference the same physically reductive properties. Indeed, it is a cogent account of *how* and *why* this conceptual dualism arose alongside a monistic physicalism that motivates all physicalist phenomenal concept strategies. Further, an explanation of this kind seems particularly pressing for the physicalist, especially in light of the intuitive anti-physicalist contention that if we endorse monistic physicalism, and uphold the claim that all concepts must be epistemologically consistent with their referent, then it seems intuitive that all conceptualisations must be as epistemologically monistic as the physically monistic properties they reference. Such that, if all concepts are referential of certain physical qualities, and these qualities are all that exist, then there should be no means of explaining how the distinguishing 'special features' of phenomenal concepts, which purportedly account for the disparity between our physical and phenomenal

conceptualisations, arose, and, as such, no means from which to explain the epistemic gap without appealing to an ontological gap.

Advocates of the PCS invoke an array of accounts to explain how monistic physicalism gives rise to this disparity in our epistemic framework. However, due to the limited scope of this thesis I intend to focus upon the most widely held of these, namely, the recognitional (see Loar 1990, Carruthers & Veillet 2007) and constitutional (see Balog 2012, Papineau 2002) accounts. The first of these, Loar's (1990) recognitional strategy, attempts to explain conceptual dualism as a natural corollary of the distinction between theoretical concepts that reference the referent physically, and 'first-order' demonstrative-recognitional phenomenal concepts that directly 'recognise' certain neurological states:

'Phenomenal concepts are recognitional concepts that pick out certain internal properties; these are physical-functional properties of the brain. They are the concepts we deploy in our phenomenological reflections; and there is no good philosophical reason to deny that, odd though it may sound, the properties these conceptions phenomenologically reveal are physical-functional properties -- but not of course under physical-functional descriptions.' (Loar 1990, p. 202)

In this account, then, there is a distinction between our physical concepts, which conceptualise the mediated, second-order physical structures and functions of a referent, and phenomenal concepts, which conceptualise the same referent through an unmediated 'direct process of recognition' (Chalmers 2010, p. 182). Further, Loar (1990) maintains that these two distinct modes of conceptualisation confer equally

distinct 'modes of presentation' (p. 204), which he employs to explain the epistemic isolation of phenomenal concepts. By means of explication for this point, let us consider the concepts of C-fibres firing and phenomenal pain. Loar's contention is that the physical concept of C-fibres firing holds a distinct 'second-order' indirect mode of presentation to the 'first order' direct mode of presentation that accompanies the phenomenal concept of pain, and it is these disparities in modes of presentation that confer disparities in the way in which we come to recognise the phenomenal and physical concepts that are, at foundation, reducible to the same physical referent. So that the physical mode of presentation presents the referent as the mediated, neurological process of C-fibre stimulation, whilst the phenomenal mode of presentation presents the neurological substrate of C-fibre excitation as an unmediated, direct recognition of pain. Thus, the phenomenal mode of presentation confers no means from which to know *a priori* that there is any physical concept of C-fibre excitation to accompany their direct phenomenal concept of pain, and the physical mode of presentation confers no means from which to know *a priori* that there is any phenomenal concept of pain to accompany the physical concept of C-fibre excitation.

In line with this, Loar (1990) maintains that, as it is not true *a priori* that C-fibres firing *is* pain; we can discern no *a priori* link between our physical and phenomenal concepts[27]. So, it is this combination of our inability to link these concepts *a priori,*

[27] It is of interest to note here that Loar's (1990) argument is strikingly similar to the identity theory espoused by Smart (1959). The primary difference is that Smart (1959) rejects the notion of *direct* phenomenal concepts, and instead maintains that the mental concept of pain is nothing beyond the designator of a specific functional role, so that our concept of pain does nothing more than to pick out a specific physical-functional property, which is a property of the brain (i.e. the brain state). Thus, whilst Smart (1959) maintains that, ultimately, mental concepts are functional concepts, and employs the notion that we can use the functional concept without knowing which brain state plays the function, in order to establish the lack of an a priori link between mental and physical concepts, Loar (1990) relies

and the distinct and direct nature of phenomenal 'modes of presentation' that, according to proponents of the recognitional framework, account for the epistemic isolation of phenomenal concepts. Thus, the central move in Loar's framework is an attempt to rely upon the 'special, direct features' of phenomenal concepts to explain why 'the idea that one picks out the phenomenal quality of cramp feeling by way of a particular feeling of cramp…is hardly incompatible with holding that the phenomenal quality is a physical property' (Loar 1990, p. 205). Indeed, proponents of this strategy contend that employing distinct modes of presentation provides a sufficient means to explain how distinct concepts arise from the same referent, and provides a grounding to appeal to the primacy of monistic physicalism, via explaining cases such as:

Phenomenal Pain (Y) = Neurological stimulation (X)

As, the concepts for Y and X both ultimately derive from the same referent (the neurological brain state), but simply deploy distinct modes of presentation that consequently confer distinctions in how we come to recognise said referent. So, Loar's (1990) basic thesis is that phenomenal pain (Y) does not provide any information about the fundamental physical structures of the neurological referent (X), but instead simply picks them out through unmediated, *direct* reference.

The strength of this thesis, and arguably the reason of its lasting appeal, rests within this notion that our phenomenal concepts *directly* denote brain states. As, it is this that granted Loar (1990) a framework with the potential to solve the Kripkean (1980)

---

upon the 'directness' of phenomenal concepts to establish the same thing. Thus, we witness two very similar means of avoiding any a priori link between physical and mental concepts. Smart's (1959) thesis shall be covered in more detail shortly.

argument for identity that had so plagued Smart's (1959) appeals to functionalism. In what follows, I shall provide a brief explication of these arguments, in the hope that, in so doing, I may facilitate a deeper elucidation of the strengths of Loar's recognitional account.

Smart's (1959) attempt to maintain a mind-brain type identity theory, and Kripke's subsequent objections, take foundation within Frege's (1948) thesis that the same referent can confer two distinct 'senses' that equally confer two distinct concepts, or terms, which can be used to denote the same referent. This contention is most famously captured within Frege's (1948) notion that 'Hesperus = Phosphorus' is true, even though such a conclusion may seem unintuitive to those who only know that the concept 'Phosphorus' references the red hue of a 'morning star' and the concept 'Hesperus' references the blue hue of an 'evening star' (p. 215). This confusion, as Frege highlights, arises as a result of both 'Hesperus' and 'Phosphorus' both denoting the same referent (the planet Venus) but simply connoting this referent amidst different 'senses', or, what Loar (1990) might term, different 'modes of presentation' (p. 203). So, the concept 'Hesperus' references Venus *indirectly* amidst the 'blue hue' mode of presentation that occurs upon Venus appearing in the evening, and 'Phosphorus' references Venus *indirectly* amidst the 'red hue' of the morning. Thus, as Frege (1948) points out, we glean a means from which to explain how two senses may arise from the same referent, and how these senses may '*indirectly'* (p. 212) connote two different senses, whilst simultaneously denoting the same reference.

For Smart (1959), this notion of *indirect* reference was enough to ground his functional theory of identity, in which he attempted to argue that, upon referencing

'pain', we do nothing beyond connoting a function, which subsequently denotes a brain state with that function. So, for Smart (1959), just as both Hesperus and Phosphorus denote Venus, but fail to capture the fundamental features of this referent by virtue of denoting it *indirectly* amidst the red/blue hue mode of presentation, our concept of pain denotes the underlying property of the brain which is identical to pain, but fails to connote the fundamental physical features of this neurological referent. Thus, Smart's (1959) central contention is that, upon, for example, referencing a phenomenal experience of 'pain', we are *directly* connoting the function of a neurological property, and, in so doing, are indirectly denoting the neurological bearer of this function, whilst leaving the fundamental physical nature of this property conceptually unrecognised. So, at foundation, just like in the Hesperus/Phosphorus case, these *indirect* references ultimately denote a singular referent, which, as far as Smart (1959) is concerned, must be the physical brain.

The problem with this, as Kripke's (1980) anti-physicalist objection points out, is that unlike in the Hesperus = Phosphorus case, in which both concepts 'pick out' their referent *indirectly*, phenomenal concepts must necessarily 'pick out' their referent *directly* as the very phenomenal property they are referent to, as:

'To be in the same epistemic situation that would obtain if one had a pain is to have a pain; to be in the same epistemic situation that would obtain in the absence of pain is not to have a pain…. Pain … is not picked out by one of its accidental properties; rather it is picked out by its immediate phenomenological quality…. If any phenomenon is picked out in exactly the same way that we pick out pain, then that phenomenon is pain.' (Kripke 1980, p. 152–53)

With this established, we reveal the strength of Loar's (1990) notion that phenomenal concepts 'directly' reference brain states. As, in positing this, Loar successfully avoids Kripke's (1980) convincing objection to Smart (1959), via firmly maintaining that our phenomenal concept of 'pain' does not connote anything, and instead simply *directly* denotes the neurological referent that *is* pain. Indeed, according to Loar (1990) it is the 'direct', first-person relationship between physical, cognitive brain states and physical, neurological referents, which confers the 'directness' of our phenomenal concepts and, in turn, confers the 'phenomenal mode of presentation' that accounts for the disparity between phenomenal and physical concepts. So, as distinct from Smart's (1959) thesis, Loar (1990) maintains that, upon applying the phenomenal concept of 'pain', we hold a special, *direct* relation to the neurological referent that *is* pain, which is distinct from the relation we hold to second-order, physical concepts that attempt to theoretically 'pick out' and analyse the structure of this neurological referent.

So, Loar's thesis hinges upon this distinction between mediated, second-order physical concepts, which present the referent physically via theoretically and structurally analysing the underlying physical structures of said neurological referents, and unmediated, *direct* first-order phenomenal concepts which present these neurological referents *directly* and phenomenally in a way that confers the directness of our physically constituted cognition to these neurological referents, but *does not* structurally analyse the underlying physicality of said neurological referent. Thus, it is Loar's insistence that phenomenal concepts *directly* reference neurological structures that has allowed him to simultaneously avoid the issues faced by Smart (1959), whilst

providing a befitting explanation for how the hard problem may be reduced to a difference in how our concepts come to present the same fundamentally physical, neurological referent.

With the strengths of Loar's (1990) thesis established, I shall briefly devote some time to a particular area of confusion, which has caused certain anti-physicalists (see Horgan & Tienson 2001) to misrepresent Loar's (1990) contentions, before concluding that, even upon correcting the arguments that arise as a result of said misrepresentations, Loar (1990) must still account for how phenomenal modes of presentation arise from purely neurological referents. In order to explicate such claims, I begin with the confusion that arises upon attempting to delineate what exactly is meant by the claim that the 'phenomenal mode of presentation' for 'pain' confers a phenomenal concept that *directly* references neurological structures (physically), whilst failing to reveal the fundamental physical nature of these structures (in this case C-fibre stimulation). It is this notion of phenomenal concepts, and in turn phenomenal modes of presentation, 'directly' denoting a neurological referent whilst seemingly not being reliant upon the fundamental, physical nature of this referent in order to 'directly' reference pain, which has caused a confusion amidst some philosophers of mind, and has led to certain philosophers misconstruing Loar's (1990) contentions. This is especially prevalent in Horgan & Tienson (2001), who seem to espouse that if Loar (1990) contends that phenomenal concepts reveal this phenomenal mode of pain 'directly and essentially' (p. 199), in a way that 'captures the essence' (p. 203) of 'phenomenal modes of presentation' (or phenomenal properties) as 'things-in-themselves' without ever appealing to physical properties, then the recognitional account 'deconstructs' itself:

1. *'When a phenomenal property (P) is conceived under a phenomenal concept (C), this property is conceived otherwise than as a physical-functional property.*

2. *When a phenomenal property is conceived under a phenomenal concept, this property is conceived directly, as it is in itself.*

3. *If (i) a property P is conceived, under a concept C, otherwise than as a physical-functional property, and (ii) P is conceived, under C, as it is in itself, then P is not a physical-functional property.*

   Hence,

4. *Phenomenal properties are not physical-functional properties.'*
   (Horgan & Tienson 2001, p. 310)

Thus, Horgan & Tienson (2001) argue that if a phenomenal property confers a phenomenal mode of presentation that distances our phenomenal concept from its neurological referent, so that when referencing 'pain' we conceive of the concept 'as it is in itself' purely phenomenally (as distinct from the physical referent of C-fibre excitation), and, under Loar's own admission, this concept confers no understanding over the physical-functional structures of the brain, then is it not intuitive to conclude that the physical-functional referent is superfluous to an explanation of these phenomenal properties, and, in turn, conclude that phenomenal properties encompass a distinct phenomenal ontology which, crucially, is *not* reducible to 'physical-functional' properties? The problem with this account, as Sundstrom (2011) formulates, is that Horgan & Tienson (2011) seem to misrepresent Loar's (1990/9) framework in their formulation of premise 2, as it is not clear that 'capturing the essence' of something necessarily confers revealing its inner nature as a 'thing-in-itself', indeed, as Sundstrom (2011) highlights, Loar explicitly distinguishes between two opposing uses of 'capture the essence of':

'On one use, it expresses a referential notion that comes to no more than 'directly rigidly designate'. On the other, it means something like 'be conceptually interderivable with some theoretical predicate that reveal the internal structure of the designated property.' (Loar 1990, p. 203)

According to Sundstrom (2011), and I would agree, 'Loar only ever acknowledges that phenomenal concepts "capture the essence" of phenomenal qualities in the former sense' (p. 276). Upon shedding this clarity unto Loar's account, we see that Horgan & Tienson (2001) fall foul to misrepresenting Loar (1990) by virtue of maintaining that phenomenal concepts 'reveal the internal structure of the designated (phenomenal) property' (Loar 1990, p. 203). This seems to be something Loar must necessarily deny in order to maintain his appeal to physicalism, and as such, phenomenal concepts must simply 'directly rigidly designate' (Loar 1990, p. 203) a phenomenal property without revealing anything about the fundamentally physical, neurological 'internal structure' of phenomenal properties.

Whilst this articulation seems to save the recognitional account from the deconstruction espoused by Horgan & Tienson (2001), I argue that it is possible to reformulate the argument of Horgan & Tienson (2001) in a manner that avoids the issue highlighted by Sundstrom (2011), by simply not committing ourselves to the claim that 'capturing the essence' of a referent should necessarily imply that we reveal its structure. In this regard, the argument could be reframed by contending that, upon maintaining that phenomenal concepts 'directly rigidly designate' their fundamentally neurological referents whilst failing to reveal the 'internal structure' of

this referent, Loar must still account for how the phenomenal modes of presentation, which are our 'direct' experiences of pain, simultaneously arise from, and distance us from, the underlying neurological structures of their referent. As, if phenomenal concepts directly refer to neurological referents 'as they are in themselves', but the phenomenal mode of presentation causes us to conceive of this referent phenomenally, in a way that hinders our capacity to recognise the fundamentally physical neurological structures of the referent, then we must ask *how* a neurologically constituted phenomenal mode of presentation (or phenomenal property) can *be* a phenomenal experience, and, in turn, must ask *how* this 'phenomenal mode of presentation' can distance us from this neurological referent, without itself being something beyond a neurological structure. As Perry (2001) formulates Max Black's argument for property dualism:

'Even if we identify experiences with brain states, there is still the question of what makes the brain state an experience, and the experience it is; it seems like that must be an additional property the brain state has…There must be a property that serves as our mode of presentation of the experience as an experience.' (p. 101)

So, in highlighting this failure to account for how this phenomenal mode of presentation arose from a purely neurological referent, this argument seems to lead the recognitional account back to the very ontological distinction it is attempting to deny. Indeed, as I shall explore later in this chapter, the need to fall back on ontologically distinct experiential properties seems to be a recurring theme for proponents of the PCS.

As such, the recognitional account seems to face a problem. However, prior to establishing any further issues with this account, it seems prudent to first elucidate a further conceptual strategy, the constitutional account `(see Papineau 2002, Balog 2012), that has the potential to provide further detail to the explanatory power of the recognitional account. Like the recognitional account, this framework employs the 'special nature' of phenomenal concepts to explain the explanatory gap, so that:

'From [the perspective of the referent constitution view], the puzzle that the explanatory gap presents is rather a trick the mind plays on itself as a result of the peculiar cognitive architecture involved in first-person phenomenal thought.' (Balog 2012, p. 31)

Thus, in essence, this account attempts to further explain Loar's (1990/9) appeals to 'phenomenal modes of presentation', and in turn explain the 'directness' of phenomenal concepts appealed to within the recognitional account, via bringing 'first-person phenomenal thought' in line with the neurological 'trick' of the mind that accompanies our phenomenal concepts. Such that, upon holding a phenomenal concept, the brain constitutes a token experience, representative of the neurologically constituted experiential type the concept is referent to, which enables us to match the referent of this concept to future token experiences of this concept. Further, it is this token experience that produces the 'phenomenal mode of presentation', or 'trick of the mind', that accounts for the 'directness' of our phenomenal concepts, and, in turn, accounts for the explanatory gap. Tye (2009) describes Balog's thesis as follows:

'*Phenomenal concept C, as used by subject S, refers to a phenomenal property or type P if, were S confronted with any token of P, she would judge that that token is a token of the same kind of experience as this token of P, where this token is an experience that S is undergoing and that thereby partly constitutes S's token of C.'* (p. 48)

So, upon employing the phenomenal concept of pain, 'an instance of the referent is *literally* (physically) present in the concept, therefore there will be always something it is like to token the concept in those applications' (Balog 2012, p. 7). Meaning that, in the course of introspecting about pain, we employ a phenomenal concept that produces a neurologically constituted token experience that represents 'what it is like' to actually undergo pain. Papineau (2002) explains this token experience in terms of an imaginative act, in which the concept of pain is constituted in the neurological states that produce the token experience of pain required to imaginatively employ pain as a referent (p. 116-120). This, according to Balog (2012), occurs because the neural constitution of the brain 'matters for reference, both in terms of reference fixing, and in terms of how the concept cognitively presents its reference' (p. 7), thus, upon remembering a phenomenal concept, the brain constitutes that concept neurologically so as to produce a 'first hand' token experience that represents the experience type 'pain' the concept references. Further, it is these neurologically constituted 'recreations' of token experiences that allow us to match the 'sameness' of one phenomenal concept to another, so that we may effectively assimilate the informational content of a referent and fix it to an appropriate phenomenal concept.

As Crane (2005) typifies, it is our neurological capacity to 'recreate, stimulate or otherwise involve referents' (p. 156) that enables Papineau and Balog to explain why

anti-physicalist philosophers often conclude in favour of ontological dualism. The argument goes that because these phenomenal concepts neurologically constitute token representations of the type of experiential states they are referent to, and physical concepts do not, then it is relatively easy to see why philosophers conclude that these concepts must be referencing distinct ontological properties. Thus, at foundation, the constitutional account just adds further explanatory weight to the recognitional account provided by Loar (1990/9), with both ultimately contending that the explanatory gap is accounted for by the 'trick of the mind' that occurs upon employing a phenomenal concept, and, crucially, maintaining that it is the special, direct nature of this 'trick' that seems to epistemologically isolate these concepts from their physical counterparts.

The problem with this strategy, as Tye (2009, p. 48) very briefly highlights, is that Balog must account for our capacity to judge between distinct experiential tokens arising from the same type of referent, and in order to do this, it seems the constitutional account would need a concept of phenomenal character. As it seems clear that in order to account for our capacity to judge the 'sameness' or distinctness of disparate tokens of experience, the constitutional account must posit a concept of phenomenal character that infers a unified experiencer, capable of judging 'what it is like' to hold different token experiences through time. This, I argue, is intuitive, as in order to identify one token experience as the same as another, this 'sameness' must be constituted in the unified, intuitive awareness of 'what it is like', so that 'phenomenal sameness becomes a sameness of phenomenal character' (Tye 2009, p. 48) that allows us to judge experiences as another of *this* or *that* kind. Without this, as Tye (2009) formulates, we would lose our capacity to demarcate different token intensities of

pain, leaving us unable to delineate between the dull, incessant pain of an ache, and the intense pain of burning. The problem with this, of course, is that as soon as the constitutional account admits reliance upon this experientially unified phenomenal character, then it seems to be positing a distinct ontology beyond that of a purely physical-functional iteration that frequently falls foul to conceptual 'tricks of the mind'.

Moreover, whilst Tye's articulation of this problem ends with his denouncement of the constitutional account, I posit that this inability to account for phenomenal character, construed as I have throughout as a unified intuitive awareness of 'what it is like', presents a deeper problem for all those who attempt to employ the PCS as a framework capable of saving physicalism from the ontological gap. The fundamental problem is that recognitional and constitutional accounts both rely upon a capacity to recognise or 'fix' referents to their appropriate concepts. However, I argue that, in taking this capacity for 'reference fixing' (Balog 2012, p. 7) as a given, the advocates of the PCS rest their framework upon the faulty premise that purely physical-functional cognitive tools necessarily confer the capacity to appropriately and causally unite a referent to a concept.

So, the fundamentals of this problem re-affirm the contentions I put forth in chapter 2 involving the difficulty of establishing concept formation without first positing a phenomenal unity capable of uniting disparate sense data experientially and *non-conceptually*. As such, my contention is that the tendency for conceptual strategies to infer that we recognise, or fix, a concept to a referent, and then neurologically recreate said referent in every instantiation of the concept, must first be predicated upon an

explanation of how we can account for the non-conceptual unification required to 'ground', and in turn recognize/conceptualize a referent, in the first place. And, as, at present, there is no convincing physical explanation for how concepts are grounded in this way, I maintain that we remain reliant upon phenomenal properties in order to solve the 'concept-grounding problem' (see Floridi 2012 and chapter 2 of this thesis for a deeper examination of this problem).

In order to effectively elucidate the intricacies of this position, let us return to the depiction of concept formation as outlined in chapter 2. This formation attempts to highlight the issue of physically explaining how we come to recognise, link, or judge referent A as corresponding to concept B without first positing an explanation for how we come to be in the state of neurological unity that allows for this link to be 'picked out'. As explicated in chapter 2, the theory that currently holds the most promise in explaining this capacity is a theory of experiential parts in conjunction with the unity thesis espoused by Chalmers & Bayne (2003), in which we may account for experiential unity by initially positing sense data A/B as experiential parts of a wider experiential whole, which subsume into a 'conjoint phenomenology for both states' (Chalmers & Bayne 2003, p. 37). Without this, a purely physical-functional depiction of the brain faces the inevitable struggle of accounting for how the informational differences necessary for the formation of concepts may be 'picked out', without either falling foul to Papineau's (1993) issue of granting the brain an already fully realised set of concepts (and thus facing the issue of explaining how concepts arose initially), or first positing a 'conjoint phenomenology' capable of unifying potentially diffuse and chaotic sense data and/or brain states.

This contention reaches to the core of my aims in chapter 2, in which I attempt to link the 'easy' problem of epistemic justification to the 'hard' problem of consciousness. Here, I maintain that the lack of a physical solution to the 'easy' problems leaves us potentially entirely reliant upon phenomenal properties in order to bring unity to an otherwise disparate and chaotic 'outer world' with the potential to confer equally disparate and chaotic sense data, and maintain that, without such properties, we are left with no convincing explanation for how the conceptualisations that underpin our 'power to judge' (McDowell 2008, p. 10) are produced. So, it seems that, without a physical explanation for both how concepts arise from purely physical properties, and why epistemic justification and conceptual formations *seem* to be experience-dependent, the explanatory success of the PCS, in the first instance of concept formation, remains reliant upon this non-conceptual phenomenal unity[28], which seems entirely irreducible to any form of 'trick of the mind', to explain our capacity to link *that* initial referent to *that* initial phenomenal concept. Further, I argue that if we accept the possibility that this initial phenomenal state 'grounds' concepts[29], then this initial state *necessarily* avoids the issues of neurological, conceptual 'tricks' or 'misrepresentations' relied upon by the PCS to explain the explanatory gap, as such tricks can only occur *after* the first phenomenal concepts have been produced by this

[28] Here I am referencing the phenomenal model of concept acquisition outlined in chapter 2 (see page 50), in which it is argued that without phenomenal properties we would (currently) be devoid of an explanation for how we achieved the capacity to form an inner locus in which referent R is matched with competing referents that are ~ R, and, subsequently, would lose both our capacity to explain how we extract (or 'pick out') information about R, as well as our capacity to explain how we form a concept to denote R. Moreover, as our current epistemic framework seems predicated upon our ability to produce conceptualisations, and as 'what it is like' sensations are epistemologically and conceptually blank (in relation to R or ~ R or indeed any referent) prior to extracting information about the nature of these referents within unified experiences, I argue that, in lieu of an adequate physical explanation for how concepts are initially formed, we are right to consider the possibility that this unified, experiential awareness of 'what it is like' is both, at least partially, epistemologically fundamental and, *initially*, entirely non-conceptual.

[29] In line with my contentions throughout chapter 2, I argue that without a physical solution to the symbol-grounding problem, the physicalist must accept the possibility that concepts are grounded phenomenally.

initial, non-conceptual experiential state. So, it seems the fundamental problem with the PCS is that without an adequate physical account of concept formation, it either becomes grounded within an, as yet unarticulated, adequate physical theory of concept acquisition (but in so doing recognises that the viability of the PCS as a solution to the 'hard' problem is entirely questionable until this future physical theory of concept acquisition *has* been articulated), or relies upon non-conceptual experience to account for the later conceptual 'tricks' that lull us into the mistaken belief that there is an ontological gap, but fails to explain how this initial non-conceptual experience, which seems entirely ontologically distinct from physical properties, arose in the first place. Ultimately, then, it seems that the extent to which we should accept the PCS is dependent upon the extent to which we accept that there will (eventually) be a physical solution to Chalmers' (1995) purportedly 'easy' problems. However, as articulated (see chapter 2), it is not, at present, convincing that such problems will ever be physically solved, and, as a result, the PCS is potentially left reliant upon, but entirely unable to physically account for, experiential qualities. As a result, it seems the PCS, as a purported solution to the 'hard' problem, cannot be taken seriously without being substantiated by an (as yet unarticulated) adequate physical theory of concept acquisition. Without this, not unlike reductive representationalism and eliminativism, we are left with a physicalist strategy that necessarily fails to convincingly incorporate experiential properties into the physicalist metaphysic.

So, as we reach the end of this chapter, and indeed the end of my examination of some of the major physicalist responses to the hard problem, we see that the two best means of reductively accounting for the ontology of phenomenal experience

(eliminativism and representationalism) fall foul to the problem of misrepresenting the problem of consciousness, and the best means of avoiding the ontological gap, via reframing it in epistemic terms, must incorporate phenomenal experience into its account of conceptual formation. Thus, whilst my account of the physical solutions to the hard problem is not exhaustive, we do witness that the ontological gap presents a metaphysical and epistemic void that some of the most championed contemporary physicalist frameworks are failing to accommodate for. In line with this, in my subsequent, and final, chapter I shall explore the most widely held contemporary anti-physicalist frameworks, which champion a shift in our metaphysic from the physicalistic monism that has predominated much of contemporary philosophy and science, to an ontological dualism that has the potential to provide a metaphysic capable of accommodating the ontology of experiential qualities.

Chapter 6

Towards a paradigm shift: supplanting monistic physicalism with ontological dualism

'*Scientists [and philosophers] are like those levers of knobs or those boulders helpfully screwed into a climbing wall. Like the wall is some cemented material made by mixing knowledge, which is a purely human construct, with reality, which we can only access through the filter of our minds. There's an important pursuit of objectivity in science [and philosophy] and nature and mathematics, but still the only way up is through the individual...In the end it's personal, as much as we want to believe it's objective.'*

(Levin 2016, p. 189)

So, like the scientist's, the philosopher's capacity to pursue truth is ultimately predicated upon a synthesis of our experiential capacity to measure reality and our capacity to produce knowledge as result of said measurements. Once this knowledge is assimilated, we begin to form a metaphysical 'wall' that provides an epistemic map of our reality, and propels the scientist forward; providing grounding from which to link one epistemic breakthrough to the next, with each breakthrough both compounding the efficacy of this metaphysical 'climbing wall', and streamlining the route to future epistemic advancement. However, as I have attempted to explicate throughout this thesis: in our creation of the wall, the contemporary mainstream have presupposed a physicalist metaphysic that, ultimately, inevitably, hinders the climb. We, amidst the naivety of our early epistemic breakthroughs, discounted the property that made the creation of the wall possible in the first place. We forgot experience.

And now our metaphysic is found lacking; its epistemic foundation crumbling under the weight of that which it can't explain. So, we now face the consequence of our failure to incorporate experiential properties into our metaphysical 'wall' at the outset. Leaving science and philosophy hindered by the early evidential successes of the metaphysical 'climbing wall' of monistic physicalism, and now unable to reconcile the ontological commitments of physicalism alongside a befitting explanation of those non-physical, experiential properties, which make enquiry possible in the first place.

And so, we reach the point of epistemic stagnation the philosophical and scientific disciplines face today, in which we confront this 'hard' problem and inevitably return to the question hinted at within the introduction of this thesis: can the experiential, observational property, which confers our means of measurement, employ purely physical-functional quantifications in order to effectively measure itself in entirety? The answer to this, in light of chapters 1, 3, 4 and 5, is, I hope, now clear. As, in order to maintain the coherency of physicalism, the physicalist arguments must persistently manipulate and misrepresent what conscious experience actually is in order to make it consistent with their metaphysic.

Thus, far from providing a coherent physicalist explanation of experiential qualities, these arguments only serve in affirming our inability to rely upon physical-functional arguments or quantifications to effectively 'measure', or in any way comprehensively explain, that which provides our means of measurement. Nor, I argue, should we expect consciousness to physically account for itself; indeed, just as a ruler cannot be employed to effectively measure itself, we should not expect ourselves, as conscious beings, to physically quantify our way out of the hard problem. So, I contend the most

obvious solution is to simply posit that physical-functional explanations, in their attempts to massage our depiction of consciousness in order to fit a physicalist framework, will never fully account for the fundamental experiential properties that are necessarily antecedent to any physical-functional theory. Thus, I ask, just as we take physical properties to be fundamental, why do we not take the experiential property that epistemologically illuminates these physical properties to be equally fundamental? This is the question that, I contend, we must pose to those advocates of physicalism, and indeed it is this question that has underpinned all my attacks on the physicalist attempts to discount the hard problem by reductively accounting for experiential properties.

Further, as I have attempted to explicate throughout this thesis, these physicalist 'solutions' consistently fail to reconcile phenomenal experience in its entirety, and, by virtue of this failure, ultimately only serve in highlighting the inability for physicalism to accommodate for the epistemic and metaphysical weight of the hard problem. As such, and as a result of the failings explicated in chapters 1, 3, 4 and 5, it seems to me that the most cogent means of escaping the metaphysical and epistemic quandary posed by the hard problem is to champion a form of dualism predicated upon the contention that both physical and experiential properties exist as ontologically fundamental components of our reality. In line with this, and in an attempt to maintain the steady march of our scientific, philosophical and, ultimately, epistemic climb, I devote this final chapter to an examination of those anti-physicalist philosophers of mind who attempt to offer metaphysical alternatives to the zeitgeist of monistic physicalism, which has so failed to accommodate for experiential properties.

Throughout this chapter, I explicate the details of the most widely held of these anti-physicalist frameworks, with a particular focus upon Swinburne's (1986-2013) substance dualism, Strawsonian (2006) panpsychism and Chalmers's (1996) notion of naturalistic dualism/ monistic property dualism. I begin with an examination of Swinburne's substance dualism, before concluding that such a thesis ultimately fails to provide a sufficiently cogent metaphysical framework. From here, I outline the virtues of panpsychism and property dualism, before arguing that monistic property dualism confers a metaphysic with the capacity to not only solve the problem of consciousness, but also be neatly reconciled alongside some of the most fundamental physical laws underpinning our understanding of reality.

6.1 Swinburne's Substance Dualism

Appeals to ontological dualism are strewn across the history of philosophical enquiry. The most widely communicated of which is the Cartesian (1641/4) notion of substance dualism, which states that our reality is constituted upon two entirely ontologically distinct entities that exist *independently* of one another: mental substances (non-physical thinking things) and physical substances (spatially extended things)[30]. Whilst this early metaphysical claim has faced fierce resistance throughout much of the contemporary philosophical literature, Swinburne (1986-2013) has very recently attempted to ignite a resurgence of this long admonished metaphysical commitment. Like Descartes (1644) before him, Swinburne (2009) maintains that as we can logically disentwine our experiential mind or 'non-physical soul' from our underlying physical body, we are right to conclude 'that each of us on Earth consists

[30] This is distinct from a neutral monism that leads to property dualism, which states that physical and mental properties are ontologically distinct, but are ultimately bound to a singular substance. This framework shall be explicated in detail shortly.

of two parts, a physical body and a non-physical soul' (p. 501). This appeal to our capacity to exist as something more than our physical body acts as the cornerstone for substance dualism, and takes foundation within the Cartesian (1644) notion that:

'While we (can) thus reject all of which we can entertain the smallest doubt, and even imagine that it is false, we easily indeed suppose that there is neither God, nor sky, nor bodies, and that we ourselves even have neither hands nor feet, nor, finally, a body; but we cannot in the same way suppose that we are not while we doubt of the truth of these things; for there is a repugnance in conceiving that what thinks does not exist at the very time when it thinks. Accordingly, the knowledge, I think, therefore I am is the first and most certain that occurs to one who philosophizes orderly.' (p. 30)

Thus, this early substance dualism was predicated upon the notion that whilst we may logically doubt that we have a body, we cannot, in the act of doubting, logically doubt that we exist, and, therefore, we must conclude that the experiential doubter, the 'I', is not a body. This argument is reflective of Descartes' essence argument, which I formulate as follows:

For any existing substance X or Y, if I can understand the essence of X without employing an understanding of the essence of Y, then I may conclude that X and Y hold distinct essences, and, therefore, conclude that X and Y are ontologically distinct substances.

Here Descartes (1644) is employing the notion of 'essence' as the 'principal property of substance…on which all others depend' (p. 240), so that this property is *essential*

to the existence of X or Y. Hence, the argument follows that if I can understand the essence of my body (Y) as a substance with the *essential* properties of being physically spatially extended, and I can understand the essence of myself (X) as an entirely distinct, experiential substance with the *essential* properties of sensation, awareness and conscious thought, then I can formulate a complete understanding of the essence of X without ever needing to appeal to the essence of Y, and, as this is possible, there must be a duality of existing substances to account for this duality in essences. Thus, Descartes maintains that as it is true that we can understand the essences of X and Y separately, then it must equally be true that bodies are in no way *essential* to the existence of the mind, as the essence of Y (the physical properties of the spatially extended body) can be logically excluded (or doubted) without ever hindering our understanding of the essence of X (the experiential properties of the non-spatial conscious mind). Indeed, as Descartes famously formulated, we can bring all things into doubt other than that which is itself *essential* to doubt (the mind). Meaning that, as the physical essence of Y can be logically excluded from our understanding of the experiential essence of X in a way that neither violates the existence of X, nor hinders our understanding of X, then we should conclude that the physical properties of the body are not *essential* properties of the mind, and therefore, should conclude that the experiential mind, the 'I', is ontologically distinct from the physical body.

Like Descartes, Swinburne (2013) upholds the notion that the only logically *essential* component of this 'I' are the mental properties essential to our mental substance, which confers our 'disposition to have sensations or thoughts or form intentions' (p. 141), and, like Descartes, maintains that a physical substance can never be logically

essential to my continued existence, as 'a physical substance [is] one for which the possession of mental properties is not essential' (p. 141). Thus, both philosophers maintain the distinction between those mental substances, which confer the mental properties that are logically *essential* to my being, and those physical substances, which confer the non-mental properties that are logically *not essential* to my being. With this distinction as their foundation, Swinburne (1986-2013) and his fellow Cartesians conclude that if my existence is logically and *essentially* predicated upon a mental substance exclusively containing mental, *non-physical* properties, then the 'I' must be capable of existing devoid of the body, as the mental substance, which is essential to my existence, is predicated upon immaterial properties, which, by their very nature, are not ontologically, or logically, dependent upon the physical properties that underpin the physical substance of the body. Hence, Swinburne and Descartes maintain that my continued existence is *essentially* constituted upon the continued existence of a mental substance, which is, in itself, *not* essentially constituted upon the continued existence of physical substances. Meaning that the soul/the mind/the 'I' must encompass a distinct mental substance that persists beyond the demise of our physicality, and therefore, must be ontologically independent of the physical body. As such, it is this Cartesian argument for the logical possibility of a mind/body separation that both underpins the core of Swinburne's (1986) work and provides the theoretical core for his resurgence of substance dualism, he lays the foundation for his thesis within his interpretation of the Cartesian framework as follows:

'The crucial point that Descartes and others were presumably trying to make is not that (in the case of men) the living body is not part of the person, but that it is not essentially, only contingently, part of the person. The body is separable from the

person and the person can continue even if the body is destroyed. Just as I continue to exist wholly and completely if you cut off my hair, so, the dualist holds, it is possible that I continue to exist if you destroy my body. The soul, by contrast, is the necessary core which must continue if I am to continue; it is the part of the person which is necessary for his continuing existence.' (p. 145)

So, as reflective of the interpretation I explicated above, Swinburne (1986) champions a reading of Descartes as a 'compound dualist' (Olson 2001, p.73), positing that whilst both the body and the soul constitute a *contingent* part of the 'I", it is only the soul (or the mind) that is truly *essential* to the 'I'[31], so that whilst one may outlive the death of the body, one may never outlive the death of the soul. Indeed, it is a compound dualism of this kind that underpins much of Swinburne's (1986-2013) appeal to substance dualism, and it is this Cartesian argument for the logical possibility of a mind/body separation that lays the foundation for Swinburne's contentions that such a separation is not just logically feasible, but also metaphysically realistic. Thus, centrally, Swinburne aims to improve upon Descartes, and instigate a resurgence of substance dualism, via showing that *if* it is logically possible to posit the mind/soul as distinct from the body, then it is logically and metaphysically *necessary* that the mind/soul[32] is ontologically distinct from the body.

Much of Swinburne's (2009/13) attempt to sharpen the link between logical possibility and metaphysical necessity takes foundation within his notion of

[31] It is of interest to note here, in line with Olson (2001), that this is distinct from 'pure dualism', which posits that whilst a physical body confers our capacity to perceive the physical world and is intimately connected to the 'I', it is not a part of the 'I'. Compound dualism, in contrast, entails that the body *is* a contingent part of the 'I', but that this part is not *essential* to the continued existence of the 'I'.

'informative designators' (Swinburne 2009, p. 512). These designators, Swinburne (2013) maintains, informatively designate a given property or substance Y so as to 'pick out' the *essential* qualities of Y, which, in turn, must rigidly designate the properties/substances that are metaphysically *necessary* to Y's *essential* existence. So, upon employing an informative designator, the existence of the referent is metaphysically necessary to the concept used to designate said referent. Meaning that, if I informatively designate the concept of 'I' so as to capture its underlying *essential* non-physical experiential essence, then the application of the concept must rigidly designate those non-physical experiential properties, which are metaphysically *necessary* in order to constitute the *essential* existence of 'I'. Swinburne uses this notion of informative designators to delineate the distinction between experiences of colour and physical wavelengths of light, and elucidate 'why physics is unable to explain how the brain-events to which the impinging light gives rise, in turn give rise to sensation of blueness' (Swinburne 1986, p. 182). He posits that, upon experiencing colour, we employ mental predicates to rigidly designate the mental properties that give rise to this experience, and these predicates are quite distinct from the physical predicates that rigidly designate physical wavelengths of light. Consequently, as both predicates are informative designators that logically capture the essential essence of their respective referents without reference to their counterpart, then we should conclude that the reason physical science fails to explain our experience of 'blueness' by appealing to wavelengths of light is because these predicates *necessarily* designate two ontologically distinct properties; one of which is the physically quantifiable wavelength, and the other is the non-physical, experiential property that is beyond the sort of thing physics deals in (Swinburne 1986, p. 182).

So, to simplify, Swinburne is maintaining that, upon informatively designating the essence of a substance Y, we simultaneously rigidly designate the existence of the metaphysical constituent that is implicit to Y. Thus, if our informative designation of wavelengths and colour experiences capture the distinct *essences* of these two referents, we must conclude that these referents contain distinct ontologies. With this established, Swinburne (2009) employs his argument for informative designation to convey that upon positing the concept of 'I', we necessarily know the fundamental nature of the referent, and, therefore, upon informatively designating the non-physical, experiential essence of this 'I', we simultaneously *necessarily* posit the metaphysical existence of the non-physical mental substance that this 'I' actually is. So that:

' 'I' or 'Richard Swinburne' as used by me… seem to be informative designators. If I know how to use these words, then… I can't be mistaken about when to apply them. My knowledge of how to use 'I', like my knowledge of how to use 'green' and 'square' means that I know the nature of what I am talking about when I use the words. Mere a priori reflection will show what my existence involves and with what it is compatible. Hence [as] there is no possibility that what I am picking out by 'I' has an underlying essence which requires me to be embodied… it follows that my existing does not involve my body existing; I am therefore a pure mental substance, essentially a soul.' (Swinburne 2009, p. 513)

Meaning, if I necessarily know the essence of 'I' upon employing the word, and, as Swinburne (2009) does, I informatively designate it as a substance containing experiential properties which confer an experiential essence that is non-physical, then

this 'I' must *necessarily* exist as an ontologically distinct, non-physical substance. Whilst Swinburne (2009/13) takes this as proof for his claim that we necessarily exist as 'pure mental substances', I posit that the immediate problem with this framework is that it seems impossible for Swinburne to maintain that his privileged epistemic access to 'I' confers a capacity to implicitly know the fundamental ontological nature of the constituents underpinning this 'I'. Indeed, if his contention is that he knows the nature of this 'I' by virtue of the mental properties employed amidst his informative designation of said 'I', then, at foundation, all Swinburne has shown is that mental properties are *essential* properties, not that the essence of this 'I' is indubitably a non-physical, mental substance. Thus, if it is only the properties themselves that are *essential* to this informative designation, then it seems equally feasible that the idea of a distinct non-physical, mental substance is superfluous to requirement, as it remains entirely possible that such mental properties may simply inhere within a singular substance, such as the brain, which seems capable of simultaneously containing physical and mental properties (this is the contention upheld by the frameworks of panpsychism and property dualism, and shall be returned to shortly).

By means of substantiating his appeal to substance dualism, Swinburne (1986) employs a number of further arguments to portray that a separation of mental and physical substances is a logically possible, and in turn metaphysically realistic, proposition. All such arguments seem to rest upon the contention that: *iff* it is logically possibility for my existence to continue even after my physical body has perished, then it is metaphysically necessary that I already contain a non-physical experiential substance, or soul, that is ontologically distinct from my physical body. In order to reach this conclusion, Swinburne (1986) employs the following argument

in an attempt to logically establish that the soul continues to exist upon the destruction of the body:

Iff:

p = "I am a conscious person and I exist in 1984."

q = "My body is destroyed in the last instant of 1984."

r = "I have a soul in 1984."

s = "I exist in 1985."

$\sim$ = Not

x = All consistent propositions compatible with (p&q) and describing 1984 state of affairs

$\Diamond$ = It is logically possible that

And:

1. p
2. $(x) \Diamond (p \,\&\, q \,\&\, x \,\&\, s)$
3. $\sim \Diamond (p \,\&\, q \,\&\, \sim r \,\&\, s)$

Then:

$p \rightarrow$ r

This can be simplified as follows:

1. I am a physically embodied conscious person and I exist in 1984.

2. It is logically possible that I exist as a physically embodied person P1 at time T1 (1984), and I remain the same existing person as P1 at T2 (1985) whilst being devoid of the physical embodiment that occurred at T1.

3. It is *only* logically possible for my existence to continue at T2 if my *essential* existence is predicated upon a non-physical mental substance, or soul; which would also imply that my existence is predicated upon this very same non-physical soul during T1 (1984).

Therefore, my conscious existence in 1984 confers my existence as a non-physical soul.

This, as combined with the argument for informative designators espoused above, represents the core of Swinburne's (1986-2013) contention that we must separate the non-physical mental substance/soul, which is both logically and metaphysically essential to the continued existence of this 'I', from the physical brain, which is *not* logically or metaphysically essential to the continued existence of said 'I'. With the core of Swinburne's appeal to substance dualism explicated, we are now in a position to address the failings inherent within positing a metaphysic of this kind.

The broadest objections to this thesis attempt to either attack the epistemic leap from conceivability to logical possibility (Zimmerman 1991), posit a violation of the laws of physics (Lycan 2002), or demand an explanation for how non-physical substances come to causally interact with their physical correlates (Kim 2009). Whilst these

arguments *do* present issues for the substance dualist, all such contentions have been objected to, and so, ultimately, the metaphysic has persisted, with contemporary proponents even positing that 'there are no good objections to substance dualism' (Gusmao 2014, p. 199). In line with this, my attack on substance dualism is predicated less upon an attempt to discount the theory entirely, and more upon an attempt to highlight its weaknesses so as to reveal the strengths of its metaphysical alternatives. Thus, I maintain that whilst substance dualism is nearly impossible to indubitably denounce, it is predicated upon a series of logical and metaphysical deficiencies which, upon being confronted, ultimately result in substance dualism giving way to a far more coherent metaphysical commitment: property dualism. In line with this, I present two issues for the substance dualism: the issue of physical/phenomenal interaction, and the issue of reconciling immaterial substances with evolutionary theory. Such problems, I argue, appear insurmountable to substance dualism, but quite compatible with a property dualism predicated solely upon the claim that physical and phenomenal properties exist as ontologically distinct properties contained within a singular neutral substance: the brain.

The first of these arguments rests upon the case for neural dependence espoused by Churchland (1984), which attempts to highlight that if substance dualism were true, there should be no evidence for neurological changes conferring changes in conscious states. This contention is explicated as follows:

'If there really is a distinct entity in which reasoning, emotion, and consciousness take place, and if that entity is dependent on the brain for nothing more than sensory experiences as input and volitional executions as output, then one would expect

reason, emotion, and consciousness to be relatively invulnerable to direct control or pathology by manipulation or damage to the brain. But in fact the exact opposite is true.' (p. 32)

Thus, Churchland (1984) highlights that, if the mind is truly an immaterial entity, then the qualities and dispositions contained within this entity should be impervious to change as a result of physical damage to the brain. This, however, as empirical science has revealed, is not the case, and, as a result, the substance dualist must account for how and why the purely physical brain may causally interact, or impact upon, the non-physical soul/entity if, as the substance dualist argues, this immaterial entity is itself wholly non-physical. Further, this problem of interaction may be inverted so as to address the difficulty of accounting for how a non-physical, immaterial entity may causally interact with the material body. In this latter case, the problem can be reduced to an explanatory issue of accounting for how a non-spatially extended mind may interact with a spatially extended body so as to be the cause of spatial phenomena, and may be articulated as follows:

1. If substance dualism is true, the non-spatial, mental properties of immaterial substances are the cause of mental events.
2. These words are spatially extended effects of mental events.
3. A spatially extended effect must be reducible to a spatially extended causal agent.
4. Thus, either substance dualism is false or epiphenomenalism[33] is true.

[33] Epiphenomenalism is the theory that mental, phenomenal properties are nothing beyond causally inert, non-physical properties that are themselves entirely dependent upon purely physical functions for causal efficacy, but contain no causal efficacy in and of themselves.

Whilst this seems to leave us entrenched in the possibility of epiphenomenal substance dualism[34], in which the substance dualist denies premise 2 on the grounds that mental events, and the mental properties that cause them, are causally inert in relation to the physical, we can further substantiate our case against substance dualism by denying the possibility of epiphenomenalism on the grounds of the 'self-stultification objection' (De Brigard 2014). The objection highlights that, in any process of justifying a belief pertaining to the causal efficacy of the mental, phenomenal properties of immaterial substances, the belief must itself be predicated upon, and caused by, the mental properties of the substance in question, and, if this is so, the epiphenomenalist, amidst their act of justifying the claim that the immaterial, mental properties of immaterial substances are causally inert, must necessarily affirm the causal efficacy of immaterial properties. This argument may be articulated as follows:

1. If epiphenomenalism is true, we hold some physically expressible knowledge about the nature of the properties of mind-like immaterial substances.
2. In order for knowledge about the properties of mind-like immaterial substances to be justified, the referent (the mental, phenomenal properties of the mind) must act as the causal agent for knowledge about mental properties.
3. If mental properties *are not* the cause of knowledge about the nature of mental properties, then beliefs about the nature of mental properties are unjustified.

[34] Epiphenomenal substance dualism maintains that, as the mind is ontologically independent from the physical body, certain mental events may cause mental events, but would maintain that this causal efficacy does not extend from the non-spatially extended mind to the spatial-extension of the physical. In this sense, the epiphenomenal substance dualist must maintain that the properties of this immaterial substance (the mind) are causally inert in relation to the physical.

4. If mental properties *are* the cause of this knowledge, then mental properties *are* causally efficacious.
5. Therefore, epiphenomenalism is either false or unjustifiable.

Hence, we can affirm both that justified, true knowledge pertaining to the nature of phenomenal properties must derive from phenomenal properties (and therefore these properties must act as the cause), and that, any attempt to establish epiphenomenalism as anything beyond a mere unjustified belief must remain reliant upon the very mental content that is itself proof for the causal efficacy of the mental properties of the mind. With this established, we can now substantiate the former argument against substance dualism by adding an additional premise and a conclusion:

1. If substance dualism is true, the non-spatial, mental properties of immaterial substances are the cause of mental events.
2. These words are spatially extended effects of mental events.
3. A spatially extended effect must be reducible to a spatially extended causal agent.
4. Thus, either substance dualism is false or epiphenomenal substance dualism is true.
5. Epiphenomenal substance dualism is not true.
6. Therefore, substance dualism is false.

As such, I argue that a coherent metaphysical framework must provide a unified theory that is capable of solving the 'hard' problem, whilst avoiding the issues of non-physical/physical causal interaction faced by substance dualism. This notion of a need

for such a theory is compounded upon attempting to reconcile substance dualism with our understanding of evolutionary theory. The problem is that it seems nearly impossible to establish how the process of Darwinian natural selection could accommodate for, or give rise to, an immaterial substance. As Lycan (2013) points out: 'nothing immaterial could possibly be adaptive' (p. 541). So, it seems Swinburne (1986-2013) is, at foundation, asserting a metaphysic that completely foregoes any attempt at integration with the theories that underpin our understanding of the natural world. As such, I posit that the ultimate problem with substance dualism is that it distances us so far from the physical sciences that it becomes incredibly difficult to reconcile an immaterial soul with our current understanding of our reality, and, as such, instead of providing coherent solutions, substance dualism leaves us with more questions than answers. In line with this, I maintain that any successful formation of ontological dualism must successfully integrate itself into our understanding of the physical frameworks underpinning the natural sciences, so as to only embolden our understanding, and mend the epistemic and metaphysical gaps, in a way that does not contravene our otherwise robust worldview. As Nagel (2012) posits:

'Even if we conclude that the materialist account of ourselves is incomplete - including its development through evolutionary theory - it remains the case that we are products of the long history the universe since the big bang, descended from bacteria over billions of years of natural selection. That Is part of the true external understanding of ourselves. The question is how we can combine it with the other things we know – including the forms of reason on which that conclusion itself is based – in a worldview that does not undermine itself.' (p. 30)

.

As such, in the final section of this thesis, I offer an analysis of two metaphysical commitments with the potential to mend the gaps in our epistemic and metaphysical frameworks in a manner that does not completely undermine either our understanding of causation, or our current worldview.

6.2 Monistic Property Dualism and Panpsychism: towards a potential solution to the Hard Problem

Property dualism is the metaphysical claim that our reality is constituted upon two distinct properties: non-physical, mental properties and physical, non-mental properties. Here, in line with Goff (2011), I employ the notion of a phenomenal, or mental, property as the experiential property 'of being a thing such that there's something that it's like to be that thing' (p. 1), and, in line with Chalmers (2003), I posit physical properties as physical structures or functions, such as mass, size, shape etc, which are devoid of an 'inner life'. With this established, this metaphysic can, in its most broad form, be explicated as follows:

'Fundamental property dualism regards conscious mental properties as basic constituents of reality on a par with fundamental physical properties, [such that] the existence of conscious mental properties is not ontologically dependent upon, nor derived from, any other properties.' (Snider 2017, p. 59)

Whilst all instantiations of property dualism adhere to this fundamental divide, I note, in line with Macpherson (2006), that the metaphysic can exhibit two distinct forms. I

reference these as type A and type B property dualism[35]. The first of which, type A property dualism, is arguably the sort held by Swinburne (1986-2013), and necessarily leads us to substance dualism via strictly maintaining that two distinct properties cannot inhere within the same substance, such that if we have an ontological duality of properties, we must have a duality in substances. The second form, type B, termed Fundamental Property Dualism, Naturalistic Dualism or Monistic Property Dualism (see Chalmers 1996), maintains a substance monism, whilst positing that physical and mental properties may simultaneously inhere within *some* iterations of this singular substance, so that the brain, for example, possesses both physical and mental properties whilst remaining a singular substance. Unlike the former iteration of type A property dualism, which inevitably seems to fall foul to the issues faced by substance dualism espoused in the previous section, this latter iteration (type B) seems to hold the potential to offer a unified solution to the hard problem, capable of, at least partially, avoiding the problems of causal interaction and the lack of a consistent worldview plaguing the substance dualist. Indeed, the virtue of monistic property dualism over its substance dualistic counterpart, which necessarily commits itself to the claim that mental properties are not spatially extended and so faces the issue of explaining how non-spatial substances may causally interact with their spatial counterparts, is that monistic property dualism appeals to a substance monism. This means that, whilst, according to the substance dualist, mental properties must be reducible to a non-physical substance that is entirely devoid of spatial extension, the property dualist may easily assert that phenomenal properties are spatially extended in the brain. This, I argue, gives us good reason to accept type B monistic property dualism over its type A substance dualistic counterpart, as upon

[35] Macpherson (2006) offers an excellent articulation of the divide between a property dualism that leads to substance dualism and a property dualism that does not.

locating phenomenal properties within the substance of the brain, we immediately grant them the spatial extension required to avoid the most pressing argument for causal interactionism erected against the substance dualist. However, whilst this does seem to present a viable reason to endorse type B property dualism over substance dualism, the objection may still be raised that the problem of accounting for precisely *how* phenomenal properties cause spatially extended effects remains unsolved. To this, in line with Chalmers (2010), I would maintain that the lack of a potential solution does not infer the impossibility of a potential solution. Indeed, just as 'Newtonian science reveals no causal nexus by which gravitation works' (Chalmers 2010, p. 126) at present, this does not imply that the causal nexus shall forever remain unknown to us, and, in line with this, we may infer that future research may provide an explanation for phenomenal-physical interaction. Thus, I argue that as monistic property dualism is both more resistant to the causal interaction objection than substance dualism by virtue of being able to account for spatial extension, and is not readily refuted by empirical objections, we have good reason to reject the type A property dualism that leads to substance dualism, and, instead, consider type B monistic property dualism as a potentially viable solution to the hard problem.

In a not wholly dissimilar vein to B property dualism[36], panpsychism, or panexperientialism, holds that phenomenal properties, in some form, exist ubiquitously so as to inhere in *all* iterations of the singular substance upon which the universe is constituted. Like property dualism, this metaphysic may be demarcated into three distinct types. Type A, constitutive panpsychism, holds that that whilst our reality is constituted upon only one substance, micro-phenomenal properties inhere in

[36] As shall be addressed later in this chapter, any similarities between the two are ultimately superficial.

all iterations of this substance, such that the macro-properties involved in animal and human consciousness are constituted upon the fundamental micro-phenomenal properties upon which the universe is constituted. This iteration may be juxtaposed alongside Type B, or non-constitutive panpsychism, in which the macro-properties of human and animal consciousness are not constituted upon any more fundamental micro-phenomenal properties, but are themselves examples of fundamental properties; as well as Type C panpsychism, or panprotopsychism (see Coleman 2016), which holds that only proto-phenomenal properties (properties that may ground phenomenal experience but are themselves not experiential in the same way as a phenomenal property) are fundamental. Thus, whilst property dualism maintains that the universe may be split into two opposing, but equally fundamental, properties, the panpsychist aims to extend the role of the physical so as to accommodate for more than the 'narrow' notion of physicality which can only account for the properties of spatial extension or function, and, instead, espouse a 'broad' notion of physicality that can accommodate for micro-phenomenal properties by maintaining that certain micro-physical properties *are* micro-phenomenal properties. Chalmers (2013) attempts to explain this by inferring that such physical properties may be demarcated as 'quiddities' that may account for 'some phenomenal properties' (p. 9), and suggests that 'perhaps the property that plays the mass role is a certain phenomenal property (or better, as mass is really a quantity) the quantity that plays the mass role is a certain phenomenal property' (p. 9). In this regard, panpsychism seems to avoid the problem of causal interaction even more readily than monistic property dualism by virtue of this assertion that phenomenal properties are physical.

Whilst this depiction of panpsychism seems to be prima facie coherent, and certainly holds well for both the non-constitutive (type B) and constitutive accounts (type A), this notion of phenomenal properties as physical properties requires a rigid definition of what a phenomenal property is, such that, in this case, we are discussing an experiential quality that may cohere with physical qualities so as to form Chalmers' (2013) 'quiddities' (p. 9). However, type C panprotopsychism fails to provide such a rigid definition by virtue of maintaining that protophenomenal properties simply exist as a 'potentiality': a propensity to produce, or ground, experiential properties, but withholding no experiential qualities in and of themselves. This is troublesome for the panprotopsychist, as it seems to produce an ambiguity regarding the nature of panprotophenomenal properties that renders panprotopyschism, in its current form, beyond the scope of coherent inquiry. Indeed, without a more robust definition of what protophenomenal properties are, we are left either confronting the possibility that we may not hold the epistemic tools to delineate the nature of proto-phenomenal properties (see McGinn's 1989 mysterianism in section 1.2 of this thesis), classifying panprotopsychism as indistinct from a form of reductive physicalism (Stoljar 2010), or halting our enquiry into panprotopyshism until a more robust definition of the nature of these non-experiential, yet not wholly physical, protophenomenal properties is articulated. As a result of this difficulty, for my purposes in this thesis, I shall be placing impetus unto the notions of Type A and B panpsychism, whilst avoiding any further detailed explication, or mention of, panprotopsychism.

As such, in line with my avoidance of both Type A property dualism and Type C panpsychism, I devote what remains of this final section exclusively to a discussion of Type B monistic/naturalised property dualism, and Type A/B constitutive/non-

constitutive panpsychism (which, due to their implicit similarities, I shall now be referencing simply as 'panpsychism'[37]). I begin with an outline of monistic property dualism, and the inevitable similarities with panpsychism, before maintaining, contrary to the contemporary arguments put forth by Goff (2011), that the panpsychist (of either a constitutive or non-constitutive inclination) must endorse a monistic property dualism in order to advance a solution to the hard problem that 'fits well with our ordinary picture of the world, as well as with the scientific picture' (Benovsky 2015, p. 335). Thus, as distinct from Goff (2011), I not only maintain that monistic property dualism holds a metaphysical weight that is unmatched by panpsychism, but that, upon endorsing a property dualism of this kind, we glean a metaphysic that both solves the hard problem and neatly integrates itself into our worldview.

Both monistic property dualism and panpsychism arose as attempts to mend the explanatory gaps inherent within physicalism and substance dualism by positing that whilst our reality is ultimately only constituted upon one type of substance, this substance holds the potential to contain experiential properties that are beyond reduction to physical properties, and, therefore, both these metaphysical commitments immediately circumvent the 'hard problem' by taking experiential properties to be fundamental constituents of our known reality. Indeed, the strength of both these metaphysical commitments implicitly rests in this appeal to establish experiential properties as ontologically fundamental constituents of a reality that, at foundation,

[37] For my purposes, a more detailed examination of the disparities between non-constitutive/constitutive panpsychism is unnecessary. Instead, as both forms are discussing (some form of) experiential quality, I deem it acceptable to simply henceforth reference panpsychism as a metaphysic that posits the ubiquitous existence of experiential qualities. Further, I legitimize this lack of detail later in the chapter, because, as I shall show, *any* experiential quality, when posited ubiquitously, creates a problem for our understanding of thermodynamics.

only contains one substance, and it is this that enables us to reconcile physical properties alongside phenomenal experience without needing to posit ourselves as a substance that is entirely distinct from the brain, or warp our depiction of what phenomenal experience actually is in order to uphold a form of reductivism consistent with physicalism. As such, these metaphysical frameworks provide a neat means from which to save physicalism from the intractable problem of establishing how experience emerged from the brain, whilst avoiding the unsavoury case for a complete separation of mind/brain entailed within substance dualism, which, ultimately, only leaves us with an entirely fractured worldview. Thus, whilst substance dualism seems to distance us from an understanding of what we are by demanding that the 'I' exists as a corporeal entity wholly distinct from the body, and physicalism seems to entirely deny what we are by warping our depiction of what phenomenal experience is, property dualism/panpsychism posit phenomenal experience, which most fundamentally constitutes 'what we are', as fundamental, and attempts to reconcile this with our understanding of physical reality so as to leave us with a metaphysic capable of solving the hard problem whilst remaining consistent with our need for a scientifically consistent, unified worldview.

Whilst I argue that, superficially, both these metaphysical commitments present viable solutions to the hard problem, my central contention throughout this final section shall be that the extent to which either metaphysic may be accepted is predicated upon the extent to which either metaphysic is capable of positing extra phenomenal properties in a manner that is consistent with our wider understanding of reality. In line with this, I posit that whilst both monistic property dualism and panpsychism appear highly similar, they remain predicated upon a variety of subtle differences, which result in

varying degrees of explanatory strength. Thus, in what follows, I attempt to delineate the fundamental differences between these arguably very similar frameworks. I begin with an explication of monistic property dualism.

Monistic property dualism can be best explicated as follows: ontologically fundamental phenomenal properties exist alongside ontologically fundamental physical properties, and both properties inhere within one type of neutral substance, which is, in itself, neither wholly physical nor mental. Thus, this thesis maintains a substance monism whilst upholding the claim that 'mental properties are fundamental properties of reality, …on a par with length, mass, electric charge, and other fundamental properties' (Churchland 1984, p. 20). So that our brain, for example, contains both non-physical, phenomenally experiential properties, which confer our experiential capacity to know 'what it is like', and physical properties, which confer the physical 'structures and functions' quantified by the physical sciences. The primary advantage of this metaphysic lies within its potential to reconcile ontologically fundamental phenomenological properties within a monistic substance replete with physical properties, as it is this that provides us with a means from which to expand our metaphysic so that we may solve the hard problem, whilst also remaining consistent with a broadly naturalistic worldview predicated upon the notion that there is only one type of substance. Indeed, it is an attempt to solve the hard problem by 'naturalising' phenomenology in this way that underpins Chalmers's (1996) appeal to 'Naturalistic Dualism'. In line with monistic property dualism, this framework aims to maintain an appeal to substance monism, whilst providing a non-reductive explanation for phenomenal experience by positing an 'extra [ontological] ingredient (p. 10) that will 'add new principles to [but not violate] the furniture of the

basic laws of nature' (p. 14), and enable a 'bottom up' explanation for phenomenal experience by positing fundamental experiential properties as constituents of our reality (p. 14). To achieve this, Chalmers (1996) predicates his metaphysic upon the notion of informational properties inhering within substances so as to create a dual-aspect phenomenon, with phenomenal properties inferring internal informational states, and physical properties inferring external informational states. This is explicated as follows:

'The ontology that this leads us to might truly be called a double-aspect ontology. Physics requires information states but cares only about their relations, not their intrinsic nature; phenomenology requires information states but cares only about the intrinsic nature. This view postulates a single basic set of information states unifying the two. We might say that internal aspects of these states are phenomenal, and the external aspects are physical. Or as a slogan: Experience is information from the inside, physics is information from the outside.' (Chalmers 1996, p. 305)

So, for Chalmers, this 'information state' is the basic state of the neutral monistic substances underlying all of reality, and this state gives rise to both phenomenal and physical informational properties. This dual-aspect theory of information underpins the entirety of Chalmers's metaphysic, and it is this appeal to the primacy of information that enables Chalmers to unify the otherwise entirely disparate ontologies of phenomenal and physical properties. As, upon positing information as the neutrally monistic fundamental state of our reality, and positing phenomenal and physical properties as derivate from this initial state, Chalmers gleans a means from which to establish an ontological link between those phenomenal-informational

properties, which inhere within a substance so as to produce internal, phenomenally informational states, and those physical-informational properties, which inhere within a substance so as to produce external, spatially extended informational states. In this regard, Chalmers (1996) presents an elegantly simple notion of monistic property dualism, as the brain becomes the informational substance in which these two properties inhere, with physical informational properties conferring the external, physicality of this substance, and phenomenal informational properties conferring this substance's inner experiential awareness of 'what it is like'.

As such, the informational substance of the brain unifies and contains these two ontologically disparate informational properties, and thus holds the potential to solve the hard problem, and account for the interaction between phenomenal/physical properties, whilst also remaining consistent with, and simultaneously expanding upon, the physical laws of nature. Therefore, this iteration of property dualism is, I argue, highly intuitive, as upon positing an extra phenomenal-informational property of this kind, we glean a means from which to explain property dualistic physical substances, such as the brain, in which dualistic phenomenal and physical properties seem to inhere, whilst also remaining consistent with our 'common sense' understanding that property monistic substances, such as chairs, exist, in which only one type of (physical) property seems to inhere. Indeed, the point to be laboured in this iteration, as distinct from the panpsychism I shall explore shortly, is that naturalised property dualism is non-committal on the extent to which phenomenal properties are ubiquitous, meaning all this property dualist upholds is that whilst our reality is constituted upon only type of informational substance, which, in order to be physically quantified by the natural sciences must necessarily contain physical

properties, *some* of these substances *may* contain further phenomenal properties. Thus, Chalmers (1996) is arguably successful in his attempt to naturalise phenomenology, as his notion of phenomenal-informational properties inhering within certain monistic informational substances seems to provide a metaphysic with the means of fully accounting for the intrinsic, internal nature of phenomenal-informational states, whilst also remaining entirely compatible with a worldview constituted upon physical, natural laws. As Chalmers (1996) himself iterates:

'The world still consists in a network of fundamental properties related by basic laws, and everything is to be ultimately explained in those terms. All that has happened is that the inventory of properties and laws has been expanded [beyond the physical properties and laws].' (p. 127)

As such, monistic property dualism, and Chalmers's (1996) Naturalistic Dualism/Neutral Monism, both simply maintain that: 1. The fundamental laws of nature are physical and a coherent metaphysic must remain consistent with said laws, and 2. physicalism is incapable of accounting for experience. Therefore, this metaphysical framework upholds the primacy of our physical sciences and natural laws by maintaining that physical properties inhere within only one type of substance, whilst also positing an extra phenomenal property that may inhere within *some* of these substances, so as to produce a property dualistic conception of reality that is capable of solving the hard problem without undermining the physical laws underpinning our worldview. Hence, Naturalistic Dualism/ Monistic Property Dualism seem to hold the potential to avoid the issues plaguing substance dualism and physicalism, via simply maintaining that whilst only one type of substance exists,

there are two types of properties (physical and phenomenal) that may inhere within this substance. However, at this stage, the extent to which this metaphysic should be championed over panpsychism remains unclear, and, as a result, I turn to an explication of panpsychism so to provide grounding from which to contrast the two metaphysics.

Panpsychism is the claim that phenomenal experience is a ubiquitous property of nature. This metaphysic is championed by a number of contemporary philosophers and scientists (see Skrbina 2007, Goff 2011, Bohm 1990, Rosenberg 2004), however, due to the limited scope of this thesis, I devote much of my discussion to the Strawsonian (2006) notion of panpsychism, which I take to be a strong elucidation of the broader panpsychist movement. Simons (2006) provides us with an effective formulation of Strawsonian panpsychism as follows[39]:

1. *'We cannot deny the existence of experience.*
2. *Experience appears to emerge from physical phenomena that are not themselves experiential*
3. *Wholly non-experiential phenomena are not by their physical nature capable of giving rise to experience*
4. *Therefore either experiences emerges magically from wholly non-experiential phenomena or the physical phenomenal from which experience emerge are in some way themselves experiential*
5. *Magical or brute emergence is absurd*

[39] I note here that Strawson (2006) himself attempts to liken his panpsychism to a form of physicalism by maintaining that phenomenal properties are encompassed within a broad 'physiCAL' notion of the physical. This has led certain commentators (see Macpherson 2006) to question whether Strawsonian panpsychism is closer to Russelian panprotopsychism, constitutive panpsychism or property dualism. However, my reading, and the one I shall be employing throughout this section, is that Strawson is a constitutive panpsychist, as he simultaneously appeals to a broad definition of the physical, whilst employing a strict definition of phenomenal properties as holding experiential qualities. Indeed, in defining phenomenal properties as such, Strawson is necessarily distanced from the vagaries of the panprotopsychist movement, which would explicitly deny the claim that 'all phenomena are in some way experiential' (Strawson 2006, p. 146).

6. *Therefore the physical phenomena from which experience emerge are in some way themselves experiential (micropsychism)*
7. *It is implausible to suppose that nature is so fragmentarily constituted that some physical phenomena are experiential while others are not*
8. *Therefore all physical phenomena are in some way experiential*
9. *But all phenomena are physical (physicalism)*
10. *Therefore all phenomena are in some way experiential (panpsychism).'* (p. 146)

With this formation as our grounding, we witness the core of the panpsychist movement. Indeed, this argument typifies the contentions that have led to the contemporary resurgence of this ancient metaphysic, which, as Skrbina (2007) highlights, in its earliest iteration was espoused by the likes of Parmenides (545-460 BC) and Plato (380 BC). This movement upholds the claim that whilst all phenomena are fundamentally physical, all such phenomena are, at foundation, also in some way experiential. Thus, not unlike the property dualist, panpsychism attempts to posit an experiential property into the fabric of our reality in a manner that is consistent with our physically monistic worldview. Strawson (2006) predicates this framework upon the contention that the ultimate nature of our reality is constituted upon 'fundamental physical entities' (p. 9), so that, at foundation, 'the universe is spatio-temporal in its fundamental nature' (p. 9), and, with this established, Strawson attempts to posits experience into this picture.

His argument follows that if we employ the traditional notion that 'physical stuff, in itself, in its fundamental nature, [is] something wholly and utterly non-experiential' (Strawson 2006, p. 11), but we know that, upon combining this physical stuff into a substance akin to a brain like ours, we witness this substance regularly constituting

experiences like ours, then we must posit that, because brute emergence is false (see section 1.2 for my explication of emergence), our traditional notion of physical stuff is incorrect, and, therefore, we must posit that this physical stuff holds some further ontologically fundamental experiential property capable of accounting for our experiential qualities. Further, as this must be the case, and as there can only be one fundamental reality, this reality must be both experiential and physical all the way down, so that *all* of reality, *all* spatio-temporal physical substances, are in some way experiential (this is the fundamental claim underpinning panpsychism); thereby making the mind, in line with the contentions of the classical panpsychist's, truly ubiquitous. Skrbina (2006) exemplifies Strawson's (2006) thesis as follows:

1.'There is one ultimate reality to the universe, which encompasses all real and concrete phenomena [physical monism]
2. Mental (experiential) phenomena are a part of this monistic reality, and hence are 'physical' (as distinct from 'physicSal', i.e. reality as described by modern physics)
3. 'Radical Kind', or brute emergence is impossible, i.e. mental phenomena cannot arise from any purely non-mental stuff (which exhibits only shape-size-mass-charge-etc. phenomena).
4. Therefore, the one reality is inherently experiential.' (p. 153)

As Skrbina (2006) highlights, this approach is very similar to the 'dual aspect monism' espoused by the likes of Spinoza (1632-77): The world is constituted upon a singular substance that 'exhibits two faces' (p. 153), producing mental properties from the inside and physical properties from the outside. Skrbina (2006) goes on to mention that this dual aspect monism 'strongly urges one' to the panpsychist claim

that phenomenal properties are ubiquitous, as it seems difficult to reconcile 'how such a world could exclude mind from any part of reality' (p. 154).

Here, I highlight most starkly the distinction between monistic property dualism and panpsychism, as property dualism makes no such claim as to the ubiquitous nature of experiential properties. Thus, whilst panpsychism is predicated upon the notion that the one fundamental physical reality must be inherently experiential, and therefore, *all* physical substances must contain phenomenal properties ubiquitously; property dualism is predicated simply upon the notion that whilst two ontologically fundamental phenomenal and physical properties exist, and our known reality is ultimately constituted upon only one type of neutral substance such that we may witness brain-like substances containing experiential *and* physical properties simultaneously, it is not necessarily true that *all* iterations of this substance must contain phenomenal properties. Indeed, for the property dualist, it seems relatively simple to maintain that certain iterations of this substance contain no phenomenal properties what-so-ever[40]. In line with this, I posit that monistic property dualism immediately presents a more coherent metaphysic, as it seems highly intuitive that not all substances must *necessarily* hold experiential qualities, and highly counter-intuitive to champion the panpsychist position that *all* substances, such as chairs or rocks, are experiential. In line with this, Chalmers (1996) flirts with the idea that 'if experience is truly a fundamental property, it seems natural for it to be widespread' (p. 297), however, ultimately, he remains unconvinced by the ubiquitous nature of the mind entailed by panpsychism, and so champions the property dualistic notion that only some substances hold experiential properties:

[40] The monistic property dualist may potentially explain this by maintaining that phenomenal-informational properties only arise within systems at a certain level of organization or function.

'I do not think it is strictly accurate to say that rocks (for example) have experiences…although rocks may have experiences associated with them… Personally, I am much more confident of naturalistic dualism than I am of panpsychism.' (Chalmers 1996, p. 299)

Contrary to this attempt to espouse the virtues of monistic property dualism as distinct from panpsychism, certain contemporary philosophers of mind (see Goff 2011), have attempted to fundamentally denounce the property dualist's 'common sense' notion that some but not all 'macroscopic objects' (p. 1) are conscious, and instead attempts to show that, contrary to our common sense intuitions, all property dualists must endorse a panpsychim that posits all objects as fundamentally experiential. Goff (2011) begins his enquiry with an elucidation of the property dualist's common sense notion that only certain objects are conscious, he elucidates this as follows:

'Commonsense Assumption: Lot's wife is conscious and a pillar of salt is not conscious.' (p. 17)

The problem with this, Goff argues, is that in order for the property dualist to maintain this common sense assumption, he must employ an argument for 'phenomenal precision', meaning that the 'cut off point' between an object being conscious/not conscious must be phenomenally precise in order to avoid the property dualist relying on a 'vague notion of whether or not a given thing is conscious' (p. 2). Goff maintains that this notion of phenomenal vagueness must be avoided by the property dualist (and I would agree), as it seems consciousness is necessarily

transparent in a way that anything with consciousness must indubitably know that it is conscious without any room for vagueness. However, according to Goff (2011), if the property dualist attempts to avoid vagueness and uphold phenomenal precision by positing, for example, that a slight adjustment to a fundamental particle confers a difference in conscious, then he faces an 'implausible consequence':

'Implausible Consequence: The fundamental psycho-physical laws which specify the physical conditions nomologically sufficient for consciousness are utterly precise, in the sense that the slightest adjustment to the smallest particle can make the difference between whether or not a macroscopic object is conscious.' (p. 17)

So, Goff's central contention is that this need for phenomenal precision makes consciousness entirely arbitrary, as it reduces the distinction between conscious/not conscious to a (potentially) miniscule adjustment to an arbitrarily precise solitary particle. This, Goff argues, distances us from the plausible notion that general laws underpin consciousness, and entrenches us within the highly implausible notion that consciousness is conferred by the minutiae of particles. In order to convey the implausibility of this, he employs the notion of a 'random fact':

'Random Fact: When a blue balloon is made from three specific kinds of elastic, A, B and C, such that there is 42% of A, 38% of B, and 20% of C, and has a certain thickness, 1 9 precise to 1,000,0000,000th of a millimetre, and is blown up such that its diameter has a certain length, precise to 1,000,000,000th of a millimetre, the balloon turns pink.' (p. 18)

Goff (2011) maintains that, if such a random fact were true to the extent that it encompassed a 'basic law of nature' (p.19), then 'we would be extremely reluctant to take it as a fundamental law, and would try to find a way of explaining its obtaining in terms of more general laws, ones which did not involve such arbitrarily precise values' (p. 19). But, this is precisely the consequence entailed by the notion of phenomenal precision, and, therefore, Goff (2011) falls back on this charge of implausibility as a means from which to reject property dualism in favour of panpsychism, contending that if something is 'extremely theoretically implausible…it is rational to avoid such a hypothesis if at all possible' (p. 12). Thus, Goff concludes: 'The property dualist should forget about common sense, and embrace conscious pillars of salt' (p. 30).

The problem with this argument is twofold, firstly, under Goff's own admission, 'it is not inconceivable' (p. 12) that consciousness *is* phenomenally precise, and, secondly, as I shall show, panpsychism is a far more theoretically and scientifically implausible hypothesis than its property dualistic counterpart. In order to make this point, I begin with an elucidation of the second law of thermodynamic theory, before attempting to logically plot the contentions of panpsychism alongside this scientifically robust, 'basic law' of reality. From here, I show that panpsychism is entirely inconsistent with our understanding of equilibrium systems and thermodynamic theory, whilst monistic property dualism is not.

The second law of thermodynamics, simply iterated, dictates that 'the entropy[41] of the (physical) universe tends to a maximum' (Clausius 1867, p. 365) until thermodynamic equilibrium is reached. So that, if we were to take water at room temperature (a temperate which is at thermodynamic equilibrium with its environment), and exert energy so as to produce 'thermodynamic work'[42] capable of bringing this water to a boil (and thus changing its temperature so that it is now at disequilibrium with its environment), then, upon no longer exerting energy on this boiling water (or the water itself no longer being able to convert energy into thermodynamic work), the second law of thermodynamics dictates that we should expect the tightly compacted, 'ordered' heat energy contained within this boiled water to, over time, disperse into its environment so as to become more disordered, or more 'entropic', until it returns to a state of equilibrium.

Whilst my analogy of boiling water may seem relatively innocuous, this law is arguably the most fundamental law underpinning our understanding of the natural world; with its explanatory value extending from ourselves, as human systems working at thermodynamic disequilibrium, to the thermodynamic reactions involved in the formation of stars (and every exchange of energy in-between). So, this law is arguably the most fundamental law underpinning all of our reality, and it suggests, simply, that all physical energy becomes more entropic (more disordered) over time, so that we witness two opposing states in which physical things exist: states of (at or close to) equilibrium, in which energy is no longer being converted into thermodynamic work, and equilibrium with the wider environment is either reached

[41] Here entropy is defined as the levels of disorder in a system, as Jorgensen & Svirezhev (2004) assert: 'it is obvious that entropy is a measure of disorder in a system' (p. 10).
[42] Here thermodynamic work is used to denote the process of energy transferal or change, either within a system, or from one system to another.

or in the process of occurring (these states are the sort witnessed in what we may term to be 'lifeless' inanimate objects, such as rocks, salt etc), and states of disequilibrium, in which energy is converted into thermodynamic work so as to maintain the system in a non-equilibrium state (these are the states witnessed within ourselves, and life more broadly).[43]

Whilst this formation of the second law is incredibly brief and arguably very basic, the crucial point here, and the point to be laboured in my attempt to reconcile panpsychism/property dualism with thermodynamic theory, is that a system may remain at disequilibrium with its environment as long as the potentially entropic energy bombarding that system may be converted (or be changed) so as to produce useful (internal) thermodynamic work. Meaning that, if we take A (human beings) and B (a grain of salt), we witness that A is able to (internally) circumvent the second law by sustainably converting potentially entropic, external, disordered energy into useful (internal) energy capable of performing thermodynamic work (this is most obviously realised in the process of consuming food), whereas B (the salt) is unable to avoid the decline into a state of maximum entropy, as it contains no means from which to convert external energy into (internal) work. As Schrödinger (1967) highlights:

'*How does the living organism avoid decay [thermodynamic equilibrium]? The obvious answer is: By eating, drinking, breathing and (in the case of plants) assimilating. The technical term is metabolism. The Greek word means change or exchange.'* (p. 71)

[43] It is crucial to note here that such non-equilibrium systems do *not* violate the second law, as ultimately (internal) entropy reduction still produces an increase in (external) entropy.

So, we see that the crucial difference between life and non-life, or disequillilbrium (life) and non-life (equilibrium) is this capacity to convert (or change/exchange) potentially entropic energy into useful internal work, so that, as Munster (1970) attests: 'a system is at thermodynamic equilibrium when, in the system, no changes of state are occurring' (p. 70), whereas a system is at thermodynamic disequilibrium when internal changes *are* occurring. With this established, we are in a position to highlight the issues for panpsychism, whilst keeping in mind the crucial notion that an internal change or conversion in a system *necessarily* implies thermodynamic disequilibrium.

In Goff's formation we witness salt (a system at thermodynamic equilibrium) and a human, conscious being (a system at thermodynamic disequilibrium), and, according to panpsychists, both systems must contain (at least some form) of the same fundamentally experiential property. This means that, if this is so, we should expect that, upon being bombarded with the energy implicit within the wider environment, both systems should experience at least some form of 'what it is like' sensation that gives rise to a spontaneous conversion of this potentially entropic external energy into some form of internal, phenomenally constituted information pertaining to 'what it is like' to have undergone this confrontation with energy. Meaning that, upon both being confronted with heat energy, for example, both should (internally) convert this energy into the sort of experiential information, which confers 'what it is like' for these system to undergo heat energy, and, as both systems are necessarily (internally) changing and converting energy in this process, we must infer that both systems are performing (internal) thermodynamic work, and are, therefore, in the strictest sense,

exhibiting thermodynamic disequilibrium. Further, as panpsychism entails that every interaction with the wider environment necessarily produces this 'what it is like' sensation which converts energy into information, then we should expect to see both systems continuously (internally) changing as they produce (internal) thermodynamic work, and, as a result, we should expect to witness two systems with the potential to avoid the immediate decline into maximal entropy, and maintain a state of thermodynamic disequilibrium, by sustaining the conversion of (potentially) entropic energy into (ordered) information. However, as we know, in relation to salt, this is simply not the case. Indeed, quite contrarily, the scientific method has demarcated a system that exhibits *none* of the qualities for self-sustaining thermodynamic work or energy conversion that we associate with non-equilibrium systems, and *all* of the qualities associated with a system that has reached maximum entropy at a state of (near) changeless thermodynamic equilibrium.

Whilst this alone seems to present a contradiction for the panpsychist, we can take this argument further, because, if salt and human beings do contain the same phenomenal property, and this property goes 'all the way' down so as to permeate every atom, quark and photon underpinning our physical reality, then we should not experience the degree of thermodynamic equilibrium we do; indeed we should not bear witness to systems such as salt at all. As, if all such fundamental particles contain phenomenal properties, and all such properties are subject to the same laws of nature, then we should expect these particles, upon forming wider physical systems, to maintain themselves away from equilibrium by constantly converting potentially entropic energy into useful (internal) information capable of (useful) thermodynamic work, and, as such, we should not expect physical systems, such as salt, in which no

conversion of energy into thermodynamic work is evident, to ever occur. Thus, it seems, upon taking panpsychism to its logical conclusion, and attempting to incorporate this metaphysic into our understanding of thermodynamics, we inevitably face the issue of attempting to reconcile a ubiquitous, fundamental phenomenal property, with the potential to convert physical energy into (internal) information so as to maintain all of reality in a state of disequilibrium, alongside an account of how we come to witness innumerable physical systems at (or moving towards) a state of thermodynamic equilibrium, which, if panpsychism were true, should be a wholly improbable state of reality.

This point is strengthened by the early findings of Sziilard (1929) and Maxwell (1888), who established that upon positing an informational component into a thermodynamic system, said system is able to sustain itself away from thermodynamic equilibrium. Further, Szillard's findings have recently been corroborated by the very contemporary work of Vidrighin et al (2016), which empirically proved that, upon a system holding a capacity for information, said system is able to both convert energy into information, and, crucially, convert information into energy so as to produce (useful) thermodynamic work. Arguably, this empirically justifies my contentions, as these theorists posit information acquisition as nothing beyond a 'measurement process' (Leff & Rex 1990, p. 28), which, in some way, extracts useful information from potentially entropic energy. This, I argue, is something that a fundamental phenomenal property simply could not avoid, as every interaction with the energy of the wider environment would confer some sort of 'what it is like' sensation that seems to necessarily, in some way, 'measure' this energy and convert it into information.

Thus, centrally, I contend that if we, as human beings, fundamentally contain the same phenomenal, informational properties as physical systems such as salt, then we should expect to witness both systems constantly converting energy into information, and thus, we should see two systems maintaining themselves at a state of non-equilibrium. However, as we implicitly know that physical systems at thermodynamic equilibrium, such as salt, do *not* exhibit any of the self-sustaining, energy conversion properties implicit to non-equilibrium physical systems, such as human beings, then I contend this gives us good reason to reject panpsychism. The core of this argument can be iterated as follows:

1. Panpsychism posits that all physical systems contain a phenomenal property.
2. All systems containing phenomenal properties must convert entropic (external) energy into ordered (internal) information.
3. Any physical system capable of this conversion must be at, and capable of sustaining, thermodynamic disequilibrium.
4. If all physical systems contain this phenomenal property, then all physical systems are capable of this conversion.
5. Therefore, according to panpsychism, all physical systems must be at thermodynamic disequilibrium.
6. Only some physical systems are at thermodynamic disequilibrium.
7. Thus, Panpsychism is false.

As such, I argue we must reject panpsychism[44], and posit a form of monistic property dualism that is able to maintain coherency with our understanding of equilibrium states and thermodynamics by simply positing that only *some* systems contain fundamental phenomenal properties (indeed, I demarcate this combination of thermodynamic theory with property dualism as a potentially interesting area for future research). In championing property dualism in this way, I argue we remain entirely coherent with the fundamental laws of thermodynamics, whilst also remaining consistent with the common sense, and empirically verified, notion that we, as conscious experiential human systems, are alive and able to sustain ourselves away from equilibrium in a way physical systems, such as salt, are not.

Thus, I bring this section, and indeed this thesis, to a close, with the contention that whilst both panpsychism and monistic property dualism present viable solutions to the hard problem, in a way physicalist theories fundamentally do not, the strength of these metaphysics is ultimately predicated upon their capacity to be neatly reconciled alongside the fundamentally physical laws that underpin our reality, and, in this case, it seems there is only one metaphysic capable of such a reconciliation: monistic property dualism. This is not to say, however, that property dualism is free from

[44] I reiterate here that I am exclusively discussing type A and B panpsychism, and highlight that this particular argument is not an effective refutation for panprotopsychism. However, I do note, in line with my previous articulation of 'Type C' panpsychism' (panprotopsychism), that this metaphysic rests upon an ambiguity that, at present, hinders earnest enquiry. Thus, in order to avoid my argument against Types A and B panpsychism, one must rest upon panprotopsychism, however, in order to do this, one must first delineate a robust definition of protophenomenal properties, as without this, we are left turning to a potential solution that produces more problems than it solves. I also note here that, alternatively, the panpsychist might coherently defend their metaphysic by providing a robust solution for what I dub the 'inverted culmination problem' (see Basile 2010, p. 98-99 for a detailed elucidation of the problem), and, in so doing, must explain how micro experiences do not necessarily produce macro experiences. This would take much of the sting out of my formulation of the thermodynamic argument against panpsychism, as this particular problem seems to be most troubling for those panpsychist that uphold the possibility of ubiquitous macro experience. However, in line with Basile (2010), I note that there is currently no convincing panpsychist theory that explains how certain micro experiences avoid becoming macro experiences, and, as a result, I maintain that, without such a theory, my argument remains intact.

experiential gaps, the most pressing of which is a problem that would demand an entire thesis in and of itself in order to coherently explain and potentially solve, and requires an explanation for how disparate, experiential parts culminate so as to form the unified bed of experiential awareness we experience today (this is the culmination problem espoused by James 1895 and is equally troubling for panpsychism).

Whilst such a problem is undeniably something that must be accounted for by the property dualist, due to the limited scope of this thesis, I do not deem it necessary to elucidate this explanatory gap in any great depth here. Instead, as I have done throughout, I devote this thesis to a more general examination of the hard problem, and an examination of some contemporary anti-physicalist solutions that are arguably worthy of future research; not a detailed examination of one particular problem (although I do highlight here that a potential area of future research may look into combining thermodynamic theory and property dualism as a means from which to solve this problem). Thus, I demarcate the fundamental contention underpinning this thesis as follows: Physicalism is shown to be false by virtue of failing to solve the hard problem, and whilst monistic property dualism is not indubitably *the* solution to this problem, this metaphysic does present certain explanatory advantages over physicalism, and does, therefore, present a solid grounding for future philosophical research by typifying a neat solution to the hard problem that is wholly consistent with some of our most fundamental physical laws.

Conclusion

What are we? This is the fundamental question underpinning the entirety of this thesis: what are the metaphysical constituents that create the bed of experiential awareness we hold? My central contention throughout has been that the phenomenal, experiential awareness, which fundamentally typifies 'what we are', confers an ontological identity that is beyond physical reduction. It cannot be posited as an emergent property of entirely non-experiential matter, eliminated or posited as an illusion, it cannot be massaged so as to fit a functional-representationalist framework, nor can it be explained away as a conceptual 'trick' that occurs upon the brain referencing itself. It is, at foundation, something we cannot deny; it is the experiential capacity that typifies the core of what we are, and, as I have attempted to exemplify, it is beyond encapsulation within the scientific physicalism that so predominates the current zeitgeist of organized philosophical and scientific thought.

So, my central contention has been that far from continuing down this path of attempting to maintain a faulty metaphysic by consistently misrepresenting experience and reducing ourselves to mechanistic, mindless, deterministic automatons, we must instead capture the distinct, fundamental ontology of experiential properties, and account for experience in its entirety, by overthrowing physicalism and once more positing ourselves as mindful, free, experiential beings. In doing this, as I have attempted to argue in chapter 6, we not only liberate ourselves from the epistemic stagnation of physicalism, but also project this mental property into our reality, so as to re-discover an experiential cosmos, which is quite distinct from the mindless, deterministic machine depicted by the physicalist movement.

Due to the ambitious scope of this thesis, and the inevitable limitations that arise upon explicating the virtues of a metaphysical paradigm shift of this kind, it seems almost unavoidable that the research evidenced herein gives rise to a range of potential shortcomings, most notable of which is my inability to accommodate for the culmination problem (James 1895) as briefly espoused within the final section. This problem highlights that, for any panpsychist or property dualist movement, the anti-physicalist must accommodate for how potentially disparate, experiential properties culminate so as to form a singular, experiential conscious being. This problem is, arguably, one of the most pressing issues facing the anti-physicalist movement, and would require a thesis with a degree of specificity in order to explicate it in any great deal. As I mentioned in 6.2, a thesis of this kind is fundamentally distinct from the broader, more general elucidation of the physicalist and anti-physicalist solutions to the hard problem that has been the focus of the research evidenced herein.

However, within the same section, I did offer a very brief hint that a potential solution to this culmination problem may be found upon combining the theories of thermodynamics and property dualism. Such a combination seems intuitive; as in order to account for the culmination of disparate experiential properties, it seems we must first posit the existence of these disparate properties into a singular physical system, so as to confer shared experiences amongst these otherwise disparate properties. Such shared experience, I argue, would be easily accounted for by thermodynamic theory. As, if all such experiential properties inhere within a system so as to produce a state of non-equilibrium, and this system is being bombarded by the *same* external energy, then it seems only natural to infer that, over time, this potentially entropic energy would be converted into a highly unified, internal

informational state, in which the once disparate, disordered and entropic 'what it is like' sensations of these experiential properties have been unified into a singular 'what it is like' informational state of the kind we experience today. So, simply, just as we expect to witness highly ordered, low entropy (internal) physical states of a non-equilibrium system, I posit that so too should we expect the phenomenal, *informational* (internal) states of such a system to exhibit low entropy. Thus, despite the scope of this thesis ultimately precluding me from a detailed formulation of the culmination problem and its subsequent potential solution, I do think that a culmination of property dualism and thermodynamic theory offers an interesting means of solving this problem, and indeed, offers an interesting area for future research in the wider field of philosophy of mind.

Whilst the scope of my research inevitably gives rise to certain flaws; it is within this scope that a great many of the virtues of this thesis lie. Chief of which is my detailed elucidation and synthesis of both the purported virtues, and, ultimate failings, of those physicalist strategies which have been most championed in the contemporary philosophical literature. Amidst this synthesis, I was able to highlight the fundamental issue with those physicalist strategies, which attempt to maintain their metaphysic in light of the hard problem, and, in turn, I was able to coherently synthesize the failings of a long-established physicalist metaphysical commitment. This fundamental issue, as I have argued throughout, is one of categorically misrepresenting phenomenal experience so as to maintain a physicalist metaphysic. The problem with this is that, in this misrepresentation, the physicalist strategies that have seen the most contemporary success inevitably only serve in both distancing us from a coherent solution to the hard problem, and starkly highlighting the explanatory inadequacies of

the metaphysic that has so predominated much of the contemporary philosophical and scientific zeitgeist.

In line with these failings, in my final chapter I expanded the explanatory scope of my research by exploring the anti-physicalist alternatives, which attempt to posit phenomenal experience as a fundamental constituent of our reality. This final chapter placed impetus upon the need to reconcile any anti-physicalist metaphysical commitments with the physical laws underpinning our understanding of nature, so as to maintain a degree of coherency that does not fracture our otherwise robust worldview. As a result, I attempted to reconcile panpsychism with thermodynamic theory, before concluding that, ultimately, as we can only conclude that *some* physical systems are working in a state of non-equilibrium, we must, therefore, reject panpsychism and endorse a monistic property dualism, which maintains that only *some* physical systems contain a phenomenal property.

Thus, I have tried to present a thesis that not only synthesizes the epistemic and metaphysical issues that arise upon endorsing physicalism, but also offers an analysis of a metaphysic, monistic property dualism, which provides a means of coherently solving the mystery of 'what we are', whilst neatly integrating itself with our scientific worldview.

47, 636 words

Bibliography

Alexander, S. (1920/2012). *Space, Time and Deity*. Nabu Press.

Allen, K. (2016) *A Naïve Realist Theory of Colour*. Oxford University Press.

Allison, E. (2004) *Kant's Transcendental Idealism: An Interpretation and Defence*. Yale University Press.

Alter, T & Walter, S. (2007) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Ashwell, D. (2003) *Conceivability and Modal Error,* in Chalmers, D. (2010) *The Character of Consciousness*. Oxford University Press.

Balog, K. (2009) *Phenomenal Concepts,* in McLaughlin, B, Beckermann, A, Walter, S. (2009) *Oxford handbook in the Philosophy of Mind.* Oxford University Press, pp. 292-312.

Balog, K. (2012) *Acquaintance and the Mind-Body Problem,* in Gozzano, S & Hill, C. (2012) *New Perspectives on Type Identity: The Mental and the Physical.* Oxford University Press, pp. 16-43.

Balog, K. (2012) *In Defense of the Phenomenal Concept Strategy*. Philosophy and Phenomenological Research, 84 (1): 1-23.

Barash, D. (2003) *Dennett and the Darwinizing of Free Will.* Human Nature Review, 3: 222–225.

Basile, P (2010) *It must be true- But how can it be? Some remarks on panpsychism and mental composition*, in Basile, P, Kiverstein, J and Phemister, P. (2010) *The Metaphysics of Consciousness: Royal Institute of Philosophy Supplement 67*. Oxford University Press.

Bateson, G. (1972/2000) *Steps to an Ecology of Mind.* The University of Chicago Press.

Batty, C. (2010) *Sense and Sensibilia.* American Philosophical Quarterly, 47 (2): 103-18.

Benovsky, J. (2015) *Dual-Aspect Monism*. Philosophical Investigations, 38 (4): 335-352.

Berkeley, G. (1710/1982) *A Treatise Concerning the Principles of Human Knowledge*. Hackett Publishing Co.

Blackmore, S. (2003) *Consciousness in Meme Machines.* Journal of Consciousness Studies, 10 (4-5): 19-30.

Blamaeur, M. (2011*) Mental as Fundamental: New Perspectives on Panpsychism.* Ontos Verlag.

Blamaeur, M. (2011) *The Mental as Fundamental*. Frankfurt: Ontos.

Block, N & Stalnaker, R. (1999) *Conceptual Analysis, Dualism, and the Explanatory Gap*. Philosophical Review, 108 (1): 1-46.

Bohm, D. (1990) *A new theory of the relationship of mind and matter.* Philosophical Psychology, 3 (2): 271-286.

Broad, C. (1925/2008) *Mind and its place in nature*. Routledge.

Brook, A. (1994) *Kant and the Mind*. Cambridge University Press.

Brown, R. (2010) *Deprioritizing the A Priori Arguments Against Physicalism*. Journal of Consciousness Studies, 17: 47–69.

Buss, D & Greiling, H. (1999). *Adaptive individual differences*. Journal of Personality, 67: 209 –243.

Byrne, A. (2002) *Don't Panic: Tye's Intentionalist Theory of Consciousness*. A Field Guide to the Philosophy of Mind Symposium. Available at: http://web.mit.edu/abyrne/www/DON'TPANIC.pdf [Accessed: 03/03/2017].

Byrne, A. (2015) *The Epistemic Significance of Experience*. Philosophical Studies, 173 (4): 947-967.

Carr, D. (1998). *Phenomenology and fiction in Dennett.* International Journal of Philosophical Studies, 6 (3): 331-44.

Carruthers, P & Veillet, B. (2007) *The Phenomenal Concept Strategy.* Journal of Consciousness Studies, 14 (9-10): 212-236.

Carruthers, P. (2000). *Phenomenal Consciousness*. Cambridge University Press.

Carruthers, P. (2005) *Consciousness: Essays from a Higher-order Perspective.* Oxford University Press.

Chalmers, D & Bayne, T. (2003) *What is the Unity of Consciousness?,* in Cleermans, A. (2003) *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford University Press, pp. 23-58.

Chalmers, D. (1995) *Facing up the Problem of Consciousness*. Journal of Consciousness Studies, 2 (3): 200-19.

Chalmers, D. (1996) *The Conscious Mind: In Search of a Fundamental Theory.* Oxford University Press.

Chalmers, D. (2002) *Consciousness and its Place in Nature,* in Stitch, S & Warfield, T. (2002) *Blackwell Guide to the Philosophy of Mind*. Blackwell. pp. 102-14.

Chalmers, D. (2002) *Does Conceivability Entail Possibility?,* In Gendler (2002) *Conceivability and Possibility*. Oxford University Press, pp. 145-200.

Chalmers, D. (2004) '*The Representational Character of Experience',* in Leiter, B. (2004) *The Future for Philosophy*. Oxford University Press, pp. 153-82.

Chalmers, D. (2006) *Phenomenal Concepts and the Explanatory Gap,* in Alter, T & Water, S. (2006) *Phenomenal Concepts and Phenomenal Knowledge: new Essays on Consciousness and Physicalism*. Oxford University Press, pp. 167-95.

Chalmers, D. (2008) *Strong and Weak Emergence,* in Davies, P & Clayton, P. (2008) *The Re-Emergence of Emergence*. Oxford University Press, pp. 244-57.

Chalmers, D. (2010) *The Character of Consciousness*. Oxford University Press.

Chalmers, D. (2010). *The Singularity: A Philosophical Analysis*. Journal of Consciousness Studies, 17: 7-65.

Churchland, P. (1984/2013) *Matter and Consciousness.* MIT Press.

Clausius, R. (1867). *The Mechanical Theory of Heat: With its Applications to the Steam-Engine and to the Physical Properties of Bodies*. J. Van Voorst: London, UK.

Crane, T. (2001) *Intentional Objects*. Ratio, 14 (4): 336-349.

Crane, T. (2003) *The Intentional Structure of Consciousness,* in Jokic, A & Smith, Q. (2003) *Consciousness; New Philosophical Perspectives*. Oxford University Press, pp. 33-57.

Crane, T. (2005) *Papineau on Phenomenal Concepts*. Philosophy and Phenomenological Research, LXXI (1): 155-162.

Crane, T. (2013) *The Given,* in Schaer, J. (2013) *Mind, Reason and Being in the World: The McDowell-Dreyfus Debate*. Routledge, pp. 229-250.

Davidson, D. (1986/2006) *The Essential Davidson*. Oxford University Press.

Dawkins, R. (1982/2016) *The Extended Phenotype: The Long Reach of the Gene.* Oxford University Press.

Dennett, D. (1991) *Consciousness Explained*. Back Bay Books.

Dennett, D. (1992) *The Self as a Center of Narrative Gravity,* in Kessel, F, Cole, P & Johnson, D. (1992) *Self and Consciousness: Multiple Perspectives.* Hillsdale, NJ: Erlbaum, pp. 103-116.

Dennett, D. (1995) *The Unimagined Preposterousness of Zombies.* Journal of Consciousness Studies, 2 (4): 322–326.

Dennett, D. (1998) *Brainchildren: Essays on Designing Minds*. Bradford Books.

Dennett, D. (2001) *The Fantasy of First Person Reports*. Available at: https://ase.tufts.edu/cogstud/dennett/papers/chalmersdeb3dft.htm [Accessed: 07/06/2017].

Dennett, D. (2001) *The Zombic hunch: Extinction of an intuition?* Royal Institute of Philosophy Supplement, 48: 27-43.

Dennett, D. (2007*) Heterophenomenology Reconsidered*. Phenomenology and the Cognitive Sciences, 6: 247-270.

Dennett, D. (2016) *Illusionism as the Obvious Default Theory of Consciousness.* Journal of Consciousness Studies, 23 (11-12): 65-72.

Dennett, D. (2017) *From Bacteria to Bach and Back: The Evolution of Minds*. Allen Lane Publishing.

Descartes, R. (1641/1996) *Meditations on First Philosophy*. Cambridge University Press.

Descartes, R. (1644/2009) *Principles of Philosophy*. SMK Books.

Dretske, F. (1995) *Naturalizing the Mind*. MIT Press.

Dretske, F. (1996) *Phenomenal externalism, or if meanings ain't in the head, where are qualia?* Philosophical Issues, 7: 143–158.

Fodor, J. (2001) *The mind doesn't work that way*. MIT Press.

Frankish, K. (2016) *Illusionism as a Theory of Consciousness.* Journal of Consciousness Studies, 23 (11-12): 11-39.

Frege, G. (1948) *Sense and Reference.* The Philosophical Review, 57 (3): 209-230.

Freedman, D., Riesenhuber, M, Poggio, T, and Miller, E. (2001) *Categorical Representation of visual stimuli in the primate prefrontal cortex*. Science, 291 (5502)

Floridi, L. (2012) *The Philosophy of Information*. Oxford University Press.

Furst, M. (2008) *Why the Phenomenal Concept Strategy Cannot Save Physicalism,* in Hieke, L & Leitgeb, H. (2008) P*apers of the 31st IWS: Reduction and Elimination in Philosophy and the Sciences*. ALWS, pp. 106-108.

G, Northoff. (2014) *Minding the Brain: A guide to philosophy and Neuroscience.* Palgrave Macmillan.

Gangestad, S & Simpson, J. (1990) *Towards an evolutionary history of female sociosexual variation.* Journal of Personality, 58: 69 –96.

Garagnani, M, Wnnekers, T, and Pulvermuller, F. (2008) A neuroanatomically grounded Hebbian-learning model of attention language interactions in the human brain. European Journal of Neuroscience, 27(2).

Gibson, J. (1979/2014) *The Ecological Approach to Visual Perception: Classic Edition.* Psychology Press: Taylor & Francis Group.

Goff, P. (2011) *Property Dualists Should Be Panpsychists.* The 3rd Online Consciousness Conference. Available at:

https://consciousnessonline.wordpress.com/2011/02/16/property-dualists-should-be-panpsychists/ [Accessed: 08/08/2017].

Goff, P. (2017) Consciousness and Fundamental Reality. Oxford University Press.

Goldman, A. (1997) *Science, Publicity, and Consciousness*. Philosophy of Science, 64 (4): 525-545.

Greenberg, R. (2001) *Kant's Theory of A Priori Knowledge*. Pennsylvania University Press.

Gusmao, R. (2014) *There are no good objections to Substance Dualism*. Philosophy, 89 (2): 199-222.

Hales, S. (2008) *What Philosophy can tell you about your Cat*. Carus Publishing.

Hanna, S. (2011) *The Myth of the Given and the Grip of the Given*. Diametros, 27: 25-46.

Harding, S. (1991) *Whose Science? Whose Knowledge? Thinking from Women's Lives*. Cornell University Press.

Harman, G. (1990) *The Intrinsic Quality of Experience*. Philosophical Perspectives, 4: 31-52.

Harman, G. (1999) *Reasoning, Meaning and Mind*. Oxford University Press.

Harnad, S. (1990) *The Symbol Grounding Problem.* Physica D, 42: 335-346.

Harnad, S. (2001) *No Easy Way Out.* The Sciences, 41(2): 36-42.

Havlik, V. (2012) *Searle on Emergence*. Organon F: International Journal of Analytic Philosophy, 19 (2): 40-48.

Horgan, T. (2013) *Original Intentionality is Phenomenal Intentionality*. The Monist, 96: 232-251.

Horgan, T & Tienson, J. (2001/2008) *Deconstructing New Wave Materialism,* in Gillett, C. (2001/2008) *Physicalism and its Discontents*. Cambridge University press, pp. 307-19.

Horrigan, P. (2007) *Epistemology: An Introduction to the Philosophy of Knowledge*. iUniverse Publishing.

Hume, D. (1968) *Treatise of Human Nature*. Oxford: Clarendon Press

Humphrey, N. (2011/2016) *Soul Dust: The Magic of Consciousness*. Quercus.

Husserl, E. (1913/1983) *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: General Introduction to Pure Phenomenology.* Kluwer Academic Publishing.

Husserl, E. (1936/1970) *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy*. Northwestern University Press.

Jackson, F. (2003) *The Knowledge Argument.* Richmond Journal of Philosophy, 1 (3): 6-10.

James, W. (1895) *The Principles of Psychology.* Henry Holt.

James, W. (1896/1985) *Is life worth living?* International Journal of Ethics, 6 (1): 1-24.

Jorgenson, E & Svirezhev, Y. (2004) *Towards a Thermodynamic theory for ecological systems*. Elsevier.

Kalloniatis, M & Luu, C. (2007) *The Perception of Color,* in Kolb, H, Fernandez, E and Nelson, R. (2007) *The Organization of the Retina and Visual System*. The University of Utah. Available at: https://www.ncbi.nlm.nih.gov/books/NBK11530/ [Accessed: 10/06/2017].

Kant, I. (1781/87/2007) *The Critique of Pure Reason*. Penguin Classics.

Kellestrup, J. (2012) *Semantic Externalism.* Routledge.

Kiefer, M & Pulvermuller, F. (2012) Conceptual representations in Mind and Brain: Theoretical Developments, Current Evidence and Future Directions. Cortex, 48: 805-825

Kim, J. (2009) *Mental Causation,* in Beckermann, A & McLaughlin, B. (2009) *The Oxford Handbook of Philosophy of Mind*. Oxford University Press.

Kremers, J, Baraas, R and Marshall, N. (2016) *Human Color Vision*. Springer International Publishing: Switzerland.

Kriegel, U. (2003) *The New Mysterianism and the Theory of Cognitive Closure.* Acta Analytica, 18 (1): 177-91.

Kriegel, U. (2017) *Reductive Representationalism and Emotional Phenomenology*. Forthcoming in Midwest Studies in Philosophy, 41.

Kripke, S. (1970/1991) *Naming and Necessity.* Wiley-Blackwell Publishing.

Kripke, S. (1971) *Identity and Neccesity,* in Munitz, M. (1971) *Identity and Individuation*. New York University Press. pp. 135-164.

Kripke, S. (1980) *Naming and Necessity*. Harvard University Press.

Leff, H & Rex, A. (1990) *Maxwell's Demon: Entropy, Information, Computing*. IOP Publishing ltd.

Leibniz, G. (1714/2014) *The Monadology*, in Strickland, L. (2014) *Leibniz' Monadology: A New Translation and Guide*. Edinburgh University Press.

Levin, J. (2016) *Black Hole Blues and Other Songs from Outer Space.* Knopf Publishing Group.

Levine, J. (1983) *Materialism and Qualia: The Explanatory Gap*. Pacific Philosophical Quarterly, 64: 354-61.

Lewis, C.I. (1929) *Mind and the World Order*. London: Constable.

Livingston, P. (2013) *Phenomenal Concepts and the Problem of Acquaintance.* Journal of Consciousness Studies, 20 (5-6): 5-6.

Loar, B. (1990/2017) *Phenomenal States,* in Loar, B. (2017) *Consciousness and Meaning: Selected Essays*. Oxford University Press, pp. 195-220.

Loar, B. (1999) *'David Chalmers's The Conscious Mind'*. Philosophy and Phenomenological Research, 59: 465–472.

Locke, J. (1689/2008) An Essay Concerning Human Understanding. Oxford University Press.

Locke, J. (1690/2015) *The Collected Works of John Locke*. Pergamon Media.

Longino, H. (1990) *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.

Luft, S. (1998)"*Husserl's Phenomenological Discovery of the Natural Attitude".* Continental Philosophy Review, 31: 153-70.

Lycan, W. (1996) *Consciousness and Experience*. MIT Press.

Lycan, W. (2002) *Philosophy of Mind,* in Bunnin, N & Janes, E. (2002) *The Blackwell Companion to Philosophy.* Wiley-Blackwell, pp. 173-202.

Lycan, W. (2013) *Is Property Dualism better off than Substance Dualism*. Philosophical Studies, 164 (2): 533-42.

MacDonald, K. (1995) *Evolution, the 5-factor model, and levels of personality.* Journal of Personality, 63: 525–567.

Macpherson, F. (2006) *Property Dualism and the Merits of Solutions to the Mind-Body Problem,* in Strawson, G. (2006) *Consciousness and its place in nature: Does Physicalism entail Panpsychism?* Imprint Academic, pp. 72-90.

Martin, M. (1998) '*An eye directed outward',* in Wright, C, Smith, B and Macdonald, C. (1998) *Knowing Our Own Minds*. Oxford University Press, pp. 99-123.

Maxwell, J. (1871/2001) *A Theory of Heat*. Dover.

Maxwell, J. (1871/2001) *The Theory of Heat.* New York: Dover.

McDowell, J. (1996) *Mind and World.* Harvard University Press.

McDowell, J. (2008) *Experience, Norm and Nature*. Blackwell Publishing.

McGinn, C. (1989) *Can we Solve the Mind-Body Problem?* Mind, 98 (391): 349-66.

McGinn, C. (1995) *Consciousness and Space,* in Shear, J. (1995) *Explaining Consciousness: The Hard Problem.* MIT Press.

McGinn, C. (2000) *The Mysterious Flame: Conscious Minds in a Material World*. Basic Books Publishing.

McGrew, T & McGrew, L. (2007) *Internalism and Epistemology: The Architecture of Reason*. Routledge.

Mill, J. (1872/1973). *System of Logic*. University of Toronto Press.

Millikan, R. (1990) *Compare and contrast Dretske, Fodor and Millikan on Teleosemantics*. Philosophical Topics, 18 (2): 151-162.

Morgan, C. (1923/2008). *Emergent Evolution*. Chandra Chakravarti Press.

Munster, A. (1970) *Classical Thermodynamics*. John & Wiley Sons inc.

Nagel, T. (1974) *What it is like to be a bat?* Philosophical Review, LXXXIII: 435-50.

Nagel, T. (1986/1989) *The View From Nowhere*. Oxford University Press.

Nagel, T. (2012) *Mind and Cosmos: Why the Materialist Neo-Darwinian conception of Nature is Almost Certainly False*. Oxford University Press.

Olson, E. (2001) *A Compound of Two Substances,* in Corcoran, K. (2001) *Soul, Body and Survival*. Cornell University Press.

Overgaard, S. (2004) *Husserl and Heidegger on Being in the World*. Kluwer Academic Publishing.

Papineau, D. (1993) *Philosophical Naturalism.* Wiley-Blackwell.

Papineau, D. (2002) *Thinking About Consciousness*. Oxford University Press.

Pautz, A. (2003) *Sensory Awareness as Irreducible: From Internalist Intentionalism to Primitivism.* Available at: http://philrsss.anu.edu.au/peopledefaults/adamp/papers/sensory_awareness_as_irreducible.pdf [Accessed: 15/06/2017].

Pautz, A. (2006) *Sensory Awareness is not a Wide Physical Relation: An Empirical Argument Against Externalist Intentionalism*. Nous, 40 (2): 205-40.

Pautz, A. (2010) A *Simple View of Consciousness,* in Koons, R & Bealer, G. (2010) *The Waning of Materialism*. Oxford University Press, pp. 25-67.

Perry, J. (2001) *Knowledge, Possibility and Consciousness*. MIT Press.

Piccini, G. (2010) *How to Improve on Heterophenomenology: The Self-Measurement Methodology of First-Person Data*. Journal of Consciousness Studies, 17 (3-4): 84-106.

Plato. (380 BC/1997) *Complete Works*. Hackett Publishing Company.

Popper, K. (1972) *Objective Knowledge: An evolutionary approach*. Oxford University Press.

Pryor, J. (2005) *There is immediate justification,* in Steup, M & Sosa, E. (2005) *Contemporary Debates in Epistemology*. Blackwell, pp. 181-202.

Putnam, H. (1967) *Psychological Predicates,* in Capitan, W & Merrill, D. (1967) *Art, Mind and Religion.* University of Pittsburgh Press, pp. 37-48.

Putnam, H. (1973) *Meaning and Reference.* The Journal of Philosophy, 70 (90): 699-711.

Putnam, H. (1975) *The Meaning of 'Meaning'*. Minnesota Studies in the Philosophy of Science, 7, p. 131-193.

Quine, W. (1956) *Quantifiers and Propostional Attitudes*. Journal of Philosophy, 50 (5): 177-187.

Rorty (1979/81) *Philosophy and the Mirror of Nature*. Princeton University Press.

Rosenberg, G. (2004) *A Place for Consciousness: Probing the Deep Structure of the Natural World.* Oxford University Press.

Russell, B. (1927) *The Analysis of Matter.* London: Kegan Paul.

Ryle, G. (1949/2000) *The Concept of Mind*. Penguin Classics.

Schnur, T, Schwartz, M, Kimberg, D, Hishorn, E, Corslett, H, and Thompson, -Schill, S. (2009) *Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area*. Proceedings of the National Academy of Sciences, 106(1)

Schrödinger, E. (1967) *What is life?* Cambridge University Press.

Searle, J. (1980) *Minds, Brains and Programs.* Behavioural and Brain Sciences, 3:417-57

Searle, J. (1992) *The Rediscovery of the Mind*. MIT Press.

Sellars, W. (1963/1997) *Empiricism and the Philosophy of Mind*. Harvard University Press.

Sheldrake, R. (2009) *New Science of Life.* Icon Books Ltd.

Simons, P. (2006) *The Seeds of Experience,* in Strawson, G. (2006) *Consciousness and its place in nature: Does Physicalism entail Panpsychism?* Imprint Academic, pp. 146-51.

Skrbina, D. (2006) *Realistic Panpsychism: Commentary on Strawson*. Journal of Consciousness Studies, 13 (10-11): 151-57.

Skrbina, D. (2007) *Panpsychism in the West*. Oxford University Press

Smart, J. (1959) *Sensations and Brain Processes*. Philosophical Review, 68 (2): 141-156.

Smith, D. (2011) *Phenomenology of Conscious Thought,* in Bayne, T & Montague, M. (2011) *Cognitive Phenomenology.* Oxford University Press.

Smith, R. (2016) *Naturalism and our Knowledge of Reality: Testing Religious Truth Claims*. Routledge.

Smithies, D. (2014) *The Phenomenal Basis of Epistemic Justification,* in Kallestrup, J & Sprevak, M. (2014) *New Waves in Philosophy of Mind*. Palgrave MacMillan.

Snider, P. (2017) *The Natural Problem of Consciousness*. Walter de Gruyer: Berlin/Boston.

Spinoza, B. (1632-77/2002) *Spinoza: The Complete Works*. Hackett Publishing Co.

Strawson, G et al. (2006) *Consciousness and its place in nature: does physicalism entail panpsychism?* Imprint Academic.

Strawson, G. (1994) *Mental Reality.* MIT Press

Strawson, G. (2006) *Why Physicalism entails Panpsychism,* in Strawson, G. (2006) *Consciousness and its place in nature: Does Physicalism entail Panpsychism?* Imprint Academic, pp. 3-30.

Strawson, G. (2008) *Raw Materialism: And other Essays*. Oxford University Press.

Stoljar, D. (2010) *Physicalism (New Problems of Philosophy).* Routledge.

Sundstrom, P. (2011) Phenomenal Concepts. Philosophy Compass, 6 (4): 267-81.

Swinburne, R. (1986/1997) *The Evolution of the Soul*. Oxford University Press.

Swinburne, R. (2002/2004) *The Possibility of Life after Death,* in Cave, P & Larvor, B. (2004) *Thinking About Death.* British Humanist Association, pp. 38-42.

Swinburne, R. (2009) "*Substance dualism".* Faith and Philosophy, 26: 501-513.

Swinburne, R. (2013) *Mind, Brain and Free Will*. Oxford University Press.

Szilárd, L. (1929) *On the Decrease in Entropy in a Thermodynamic System by the Intervention of Intelligent Beings*, in Harvey, S. & Rex, A. (2014) *Maxwell's Demon: Entropy, Information, Computing*. Princeton University Press, pp. 124-34.

Tartaglia, J. (2015) *Philosophy in a Meaningless Life: A System of Nihilism, Consciousness and Reality*. Bloomsbury.

Tononi, G. (2012) *Phi: A Voyage from the Brain to the Soul*. Pantheon.

Tye, M & Byrne, A. (2006) *Qualia ain't in the Head.* Nous, 40 (2): 41-55.

Tye, M. (1995) *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press.

Tye, M. (1999) '*Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion'*. Mind, 108 (432): 705-25.

Tye, M. (2000) *Colour, Content and Consciousness*. Cambridge, Mass: MIT Press, Bradford Book.

Tye, M. (2002) Representationalism and the Transparency of Experience. Nous, 36 (1): 137-151.

Tye, M. (2009) *Consciousness Revisited: Materialism about Phenomenal Concepts*. MIT Press.

Valaris, M. (2008) *Inner Sense, Self-affection, and Temporal Consciousness in Kant's Critique of Pure Reason.* Philosophers' Imprint, 8 (4): 1-18.

Vidrighin, M, Dahlsten, O, Barbieri, M, Kim, M, Vedral, V, Walmsley, I. (2016) *Photonic Maxwell's Demon.* Physical Review Letters, 116 (5): 426-459.

Wah, T. (2007) *Heterophenomenology Debunked.* Journal of Humanities & Social Sciences, 1 (1): 1-11.

Wheeler, M. (2010) *In Defense of Extended Functionalism*, in Menary, R. (2010) *The Extended Mind*. MIT Press, pp. 245-270.

Williams, B. (1986) *The View From Nowhere: A Passion for the Beyond*. London Review of Books, 8 (14): 5-6.

Wilson, J. (2005) *Supervenience-Based Formulations of Physicalism*. Noûs, 39 (3): 426-59.

Wittgenstein, L. (1969) *On Certainty.* London: Harper & Row.

Worley, S. 2003. *Conceivability, possibility and physicalism.* Analysis 63:15-23.

Wylie, A. (2003) "*Why Standpoint Matters",* in Figueroa, R & Harding, S. (2003) *Science and Other Cultures: Issues in Philosophies of Science and Technology*. New York and London Routledge, pp. 26–48.

Wyss, P. (2004) *Two Aspects of Emergence*. Richmond Journal of Philosophy, 8: 1-9.

Wyss, P. (2012) *Emergence, Neither True Nor Brute.* Journal of Consciousness Studies, 19 (9-10): 220-236.

Yablo, S. (1993) *'Is conceivability a guide to possibility?'* Philosophy and Phenomenological Research, 53: 1–42.

Yablo, S. (2000) *"Textbook Kripkeanism and the Open Texture of Concepts".* Pacific Philosophical Quarterly, 81: 98–122.

Yablo, S. (2008) *Thoughts: Papers on Mind, Meaning and Modality*. Oxford University Press.

Zawidzki, T. (2007) *Dennett*. Oneworld Publiciations.

Zimmerman, D. (1991/2003) "*Two Cartesian Arguments for the Simplicity of the Soul",* in O'Connor, T & Robb, D. (2003) *Philosophy of Mind: Contemporary Readings*. Routledge, pp. 15-30.