



This work is protected by copyright and other intellectual property rights and duplication or sale of all or part is not permitted, except that material may be duplicated by you for research, private study, criticism/review or educational purposes. Electronic or print copies are for your own personal, non-commercial use and shall not be passed to any other individual. No quotation may be published without proper acknowledgement. For any other use, or to quote extensively from the work, permission must be obtained from the copyright holder/s.



**Neutral Monism Against the Qualiophiles**

**George Peter Carpenter**

**Doctor of Philosophy in Philosophy**

**October 2018**

**Keele University**



# Abstract

This thesis argues for an austere form of monism that incorporates aspects of panpsychism and physicalism, with the aim of putting naturalist epistemology on a secure footing. I criticise panpsychism for failing to live up to its promises of defending what we ordinarily think of as consciousness against physicalism, and criticise the metaphysical presuppositions of its current high-profile proponents. These presuppositions are contrasted with more recent approaches in philosophy of science. The mind-body problem itself endures these assaults, however, and I criticise physicalists who claim their position is the more common-sensical, along with naturalists who think they can avoid metaphysics. Both tendencies are represented by phenomenal concept strategists, whose position comes to seem extreme over the course of two chapters. I then offer my own solution to the mind-body problem. My position seeks a dialectical reconciliation between the possibility of directly experiencing reality, associated with anti-physicalist mysticism, and physical reductionism. I therefore take time to establish both the historical novelty of physicalism, and aspects of continuity which it may share with its predecessors.

*Key words: consciousness, panpsychism, physicalism, neutral monism, qualia, reductionism, philosophy of science, philosophy of mind, history of ideas*



# Contents

<b>Introduction.....</b>	<b>1</b>
<b>Chapter 1: Venturing on the Seas of Panpsychism.....</b>	<b>9</b>
<b>[1.1] Chalmers' Legacy.....</b>	<b>9</b>
<b>[1.2] Ghosts versus Zombies.....</b>	<b>14</b>
<b>[1.2.1] The Possibility of Having 'All the Facts'.....</b>	<b>14</b>
<b>[1.2.2] Varieties of A Priori Metaphysics.....</b>	<b>20</b>
<b>[1.3] Russellian Monism.....</b>	<b>25</b>
<b>[1.4] Goff in Focus: Foundations.....</b>	<b>33</b>
<b>[1.5] Goff in Focus II: Speculations.....</b>	<b>41</b>
<b>[1.5.1] Aristotle .....</b>	<b>43</b>
<b>[1.5.2] Mental Causation.....</b>	<b>46</b>

<b>Chapter 2: The Bonding Problems.....</b>	<b>61</b>
[2.1] Problems and Prospects.....	61
[2.2] Lockwood’s Quantum Consciousness.....	66
[2.2.1] Lockwood’s Mind-independent Minds.....	70
[2.2.2] The Sceptical Wedge.....	79
[2.3] Against Humean Metaphysics.....	83
[2.3.1] John Heil: Powers and Qualities.....	84
[2.3.2] Wesley Salmon on Scientific Explanation.....	89
[2.4] Wilkes on Personal Identity.....	94
[2.4.1] Panexperientialism Again.....	100
[2.5] Rosenberg’s Ontology.....	102
 <b>Chapter 3: Historicism and Naturalism.....</b>	 <b>115</b>
[3.1] Aristotle on Matter.....	115
[3.2] Aristotle on Mind.....	121
[3.3] Descartes Versus Aristotle.....	128
[3.4] Science Versus Common Sense.....	132
[3.4.1] Rosenberg’s Ambiguities.....	132
[3.4.2] Common Sense and the Principle of Separability.....	141



[3.4.3] John Heil and the Fate of Direct Realism.....	147
[3.4.4] Naturalism and Scepticism.....	156
[3.5] Functionalist and Teleological Semantics.....	162
[3.6] The Heuristic Psychology of Science.....	171
[3.7] Relative Common Sense.....	179
<b>Chapter 4: The Phenomenal Concepts Strategy.....</b>	<b>185</b>
[4.1] Overview.....	185
[4.2] Strong Necessities.....	190
[4.3] Objections.....	194
[4.3.1] Dual Revelation.....	195
[4.3.2] The Possibility of Metaphysics.....	197
[4.3.3] Translucent Consciousness.....	202
[4.3.4] Raffman Qualia.....	206
[4.4] Types of Conceivability.....	210
[4.5] The Use-Mention Distinction.....	222
[4.5.1] Unbelievable Physicalism.....	225
[4.5.2] Using versus Mentioning Heuristics.....	228

<b>Chapter 5: The Disconsolations of Physicalism.....</b>	<b>235</b>
[5.1] Is PCS Illusionist?.....	235
[5.2] Raffman and Balog on “Presentational” Concepts.....	244
[5.3] Partial Revelation and Selfhood.....	255
[5.4] Gunderson on Stepping Outside of Ourselves.....	263
<b>Chapter 6: Neutral Monism Against the Qualiophiles.....</b>	<b>273</b>
[6.1] Wimsatt’s Incoherence.....	273
[6.2] Rationalist Metaphysics and ‘Stress Free’ Modal Dualism.....	279
[6.3] Translucent Consciousness and Neutral Monism.....	286
[6.4] Globus on Pure Consciousness.....	392
[6.5] Some Epistemological Problems.....	300
[6.5.1] Scepticism.....	302
[6.5.2] Knowledge.....	310
[6.6] Conclusion.....	312
<b>References.....</b>	<b>315</b>



# Acknowledgements

Many thanks to James Tartaglia, Giuseppina D'Oro, Sorin Baiasu, Stephen Leach, Adam Balmer, Christopher Murphy, Anne-Elizabeth de Merode, Sebastian Orlander, Jonathan Head, Sam Thomas, Greg Oliver, George House, Damian Veal (RIP), and all other friends, colleagues and family – this thesis would be much the poorer without you.





# Introduction

This thesis is directed against philosophers of mind influenced by David Chalmers, who appeal to phenomenal qualities, or qualia, to ground arguments for the irreducibility of mind to matter. Chalmers' arguments, if sound, challenge both physicalism and naturalism. Physicalism is defined by Chalmers as the claim "that everything in the world is physical, or that there is nothing over and above the physical, or that the physical facts in a certain sense exhaust all the facts about the world".<sup>1</sup> It is an ontological or metaphysical thesis about the nature of the mind-independent world, and about the mind as being just as physical, or dependent upon the physical, as any other object. While naturalism is sometimes thought to be synonymous with physicalism (e.g. metaphysical naturalism), *methodological* naturalism makes a distinct claim about how we come to know, either that physicalism is true, or that alternative ontological claims at least are false.

According to this thesis, the best methods that we have for finding out about the world are the methods, whatever they are, that one finds in the natural sciences... It says that we should think about the world in the way that those people think about the world, where the demonstrative expression 'those people' picks out a particular class of people, viz. scientists. (Stoljar 2010, p11-12)

---

<sup>1</sup> Chalmers (1996), p41

This view is also associated with suspicion of *a priori* philosophical theorizing; Penelope Maddy (2007) dubs naturalism “second philosophy” (403-4) on that basis. By contrast, Chalmers and, indeed, most philosophers upholding the existence of a ‘hard problem’ of consciousness are engaged in ‘first’ philosophy of one sort or another. (Physicalism or metaphysical naturalism will be the focus of chapter 1 and chapter 4; chapters 2 and 3 deal more with methodological naturalism, and chapters 5 and 6 discuss both.)

Dennett (1991) somewhat awkwardly calls these philosophers “qualophiles” (Dennett, 1991, p386); I call them “qualiophiles”. I contend that, insofar as qualia figure in metaphysical explanations of how experience and the physical hang together, they do more harm than good, obfuscating issues that are best kept separate from the metaphysics of mind, and adding epicycles to the latter. Nevertheless, I maintain that there is a best of both worlds to be had that gives the mental something of the pride of place accorded it by Chalmers and others, while dovetailing with scientific explanation in a relatively elegant manner. As I discuss in chapter 1, in current philosophy of mind the qualiophile vanguard is now often associated with panpsychism, the view that everything in the universe has consciousness. The reason that conception of consciousness has been influential is due to its relationship with metaphysical essentialism, defended by Saul Kripke, and a position known as Humean supervenience, defended by David Lewis. These positions come in for criticism in chapter 2. Throughout this thesis, I argue that philosophers of mind have overrated the proximity of the mind-body problem to issues related to physical composition, and that panpsychists undermine themselves by construing fundamental experiential reality as almost entirely foreign to the kind of experience we ordinarily take ourselves to have. These two presuppositions are evident in Chalmers’ belief that any solution to the mind-body problem will be highly speculative, drawing on cutting-edge physics and as-yet-unknown ways of conceiving experience itself.



Chalmers' optimism that philosophy could yet solve the mind-body problem along the lines he proposes is mirrored by Phillip Goff's pessimism, who denies that an intelligible link between fundamental experience and ordinary human consciousness can be forged, and concludes that the scientific worldview must be fundamentally mistaken if the mind-body problem is to be solved at all. Goff's objection is a response to Chalmers situating experience so far outside of our capacity for self-reflective judgements about it that that we no longer seem to have introspective access to it. While not incoherent, Chalmers' radical epistemic externalism, which leads him to epiphenomenalism about consciousness, contains the germ for his brand of "constitutive" panpsychism, which for Goff is no better than physicalism. Against 'weird and wacky' micro-phenomenal properties, Goff upholds that the only viable consciousness is the holistic, rich kind that we ordinarily enjoy, and models his ontology on this insight, which he believes goes against the 'grain' of modern physics.<sup>2</sup> However, because Goff's conception of consciousness is so far removed from science, there are no grounds for identifying it with any known properties of our bodies. Indeed, as Goff develops his ideas, consciousness increasingly begins (literally and figuratively) to float free from the body or any other physical marker of personal identity and thus ultimately, once again, from phenomenal judgements.

Chapter 2 argues that the problems caused by Chalmers' and Goff's accounts of consciousness are variations of the "phenomenal bonding problem", which is essentially the mind-body problem for panpsychists. I return to the issue of phenomenal judgement, in the context of an account which both meets Chalmers' demands for a response to the bonding problem, while highlighting defects which the remainder of the thesis will argue are unavoidable, since they are just the consequences of treating consciousness as an objective,

---

<sup>2</sup> For full exposition of the jargon used here, see chapter 1.

mind-independent property: the 'essence' of matter.<sup>3</sup> I argue that the issues of phenomenal judgement and ontology need to be solved simultaneously in order for any solution to the mind-body problem to work, and in this chapter I examine several attempts to do so.

Michael Lockwood Lockwood and Gregg Rosenberg fill out the details of Chalmers and Goff's speculative proposals, and yet underscore the distance between any conception of consciousness rooted in physics and that which we actually experience. Rosenberg in particular marks a 'return to Goff', and bears out Goff's claim that conscious subjects must be determinate, fundamental entities.

Importantly, however, Rosenberg ends up with scientific anti-realism, which then undermines the motive for defending panpsychism in the first place. The motive, according to Chalmers (2015), comes from a "dialectical" argumentative strategy, in which the response from philosophers confronted with a mutually exclusive thesis and antithesis should be to seek a middle way between the two of them, rather than choosing one or the other (*ibid*, p247-8). For philosophy of mind, the thesis is dualism, and the antithesis physicalism. Panpsychism resolves the tensions between these two views by asserting the existence of a mind-independent physical world for scientists to study, as does physicalism, but upholds the existence of conceptually irreducible mental properties, like dualism – only these are properties possessed by material objects, not disembodied souls. Hence the need for panpsychists to show how we could know the nature of experience, given what we know about neuroscience and psychology, and how we could know about the external world, given its relationship with experience.

The position I defend – neutral monism, a 'watered down' form of panpsychism – seeks to show how we can know about the mind as well as the world. And I arrive at neutral monism through an extension of the dialectical methodology which motivated panpsychism

---

<sup>3</sup> I point out the tensions in that conception of mind in chapter 2, but ultimately conclude – by consideration of Diana Raffman's work on philosophy of perception – that the science of perception supports it after all.

in the first place. All the philosophers I discuss are either argued to have overlooked a crucial tension between two theses which they defend, and which must be resolved, or to exemplify an antithesis of some position which another philosopher failed to consider. Chapters 2 and 3 will argue that Chalmers and Goff's appeals to essentialism make presuppositions which no disinterested observer need accept, and contrast them to positions diametrically opposed to them in metaphysics, philosophy of science, philosophy of psychology, and even theology.

However, there is a further motive for looking at these debates from the standpoint of a disinterested observer, as I put it above. In the course of criticising some of Chalmers' argumentative strategy, Goff appeals to Cartesian scepticism as the most reliable way of motivating the existence of irreducible subjective properties. Consciousness emerges as that which is most resistant to doubt. And the emphasis on scepticism ought to motivate scepticism of positions whose conception of the nature of consciousness, and their solutions to the mind-body problem, make too many presuppositions. Indeed, the very possibility of formulating a conception of experience about which we enjoy an authoritative level of certainty, but which can be reconciled with scientific authority on all other matters of fact, including psychological facts, suggests that that conception, and the certainty that goes with it, had better be pretty minimal. Minimalism was part of the appeal of Chalmers' brand of panpsychism, and so the criticisms of him voiced by Goff and others must be addressed, since they have the effect of complicating panpsychism to the point that it is no longer a compelling alternative to physicalism or dualism.

So the dispute between Goff and other forms of panpsychism becomes the first major division which philosophers in chapter 2 are introduced to solve. But many more antinomies emerge in the course of explicating their theories. Thus, while chapter 1 introduces Chalmers and Goff as committed to modal rationalism, the view that possibilities

can be known a priori, and logical supervenience, which prizes conceptual reduction over empiricism, chapters 2 and 4 bring these positions into dispute with defenders of strong necessities, or “non-logical” modality, represented chiefly by Wesley Salmon, John Heil, Kathleen Wilkes, and Stephen Boulter. Chalmers’ and Goff’s reductionism and atomism is also put under pressure in chapter 2 not only by the aforementioned defenders of non-logical modality, who are largely naturalists, but by fellow panpsychists Michael Lockwood and Gregg Rosenberg as well.

Chapter 3, in many ways the heart of this thesis, goes further afield, and situates the tensions left outstanding at the end of chapter 2 in the context of disputes between Aristotelian and Cartesian philosophy. This chapter is preeminent because it begins with an attempt, by medieval Aristotelians, to resolve a dialectic internal to Aristotle’s metaphysics – between universals and particulars, temporal and unchanging reality, common sense and revisionist metaphysics – in a way which provides the template for my own attempts to save panpsychism in chapter 6. However, I also seek to credit the historical rupture of modern science from the earlier Medieval worldview, exemplified by Descartes, a rupture characterised by the division between direct and indirect realism. I then seek to show how naturalist epistemology, which is at the same time an attempt at scientific psychology, inherits much of Descartes’ worldview. Yet, while that worldview is, I argue, in fundamental tension with common sense, many naturalists portray themselves as saving common sense from scientific revisionism, and as enemies of Descartes (chapter 3 focuses on Boulter and William Wimsatt, but these sections build on the previous chapter’s exposition of Wilkes and Heil).

That portrayal is pressurised in chapters 4 and 5, which focus on a version of physicalism known as the phenomenal concepts strategy (PCS), which is diametrically opposed to the modal rationalist, non-naturalist metaphysics of the panpsychists surveyed in

chapters 1 and 2, along with the suggestions of illusionism – the view that consciousness is an illusion – in Wilkes and Wimsatt. PCS is portrayed by its proponents as doing minimal violence to our intuitions about mind and nature. Hence, the upshot that PCS is ultimately complicit with illusionism, and thus cannot save common sense (as I conclude in chapter 5), helps motivate a reassessment of the un-commonsensical conclusions of panpsychists discussed earlier, and to seek reconciliation between them and their opponents. I attempt to do so in chapter 6, where the dispute between naturalism and first philosophy is resolved through seeking a middle way between scepticism and realism, a ‘metaphysical’ interpretation of scepticism which yields a minimal definition of consciousness as pure, indeterminate presence, and connects this with the conception of matter advanced by Descartes. Finally, I conclude with a note of caution: might the real problem of consciousness have slipped through our fingers?



# Chapter 1:

## Venturing on the Seas of Panpsychism

### [1.1] Chalmers' Legacy

When David Chalmers' *The Conscious Mind* was published in 1996, it crystallized several decades' worth of arguments, in the analytic philosophical tradition, for construing consciousness as a unique challenge to the scientific worldview. It drew on a range of material, associated with philosophers such as Frank Jackson (1982), Thomas Nagel (1974), Joseph Levine (1983) and Saul Kripke (1980), whose contributions could subsequently be seen as part of a more or less unified camp – what Daniel Dennett calls the “B” team – opposed to the encroachment of physicalism<sup>1</sup> represented by Dennett's “A” team, indicating what he considers to be its status as the default position for most scientists.<sup>2</sup> Over two decades later, the conception of consciousness defended in *The Conscious Mind* (henceforth *TCM*) continues to set the agenda for arguments about the mind in the Anglophone world: there Chalmers posed consciousness as a truly “hard problem” for philosophers to solve.<sup>3</sup>

---

<sup>1</sup> The position is occasionally still labelled ‘materialism’; I shall treat the two as synonymous unless otherwise noted.

<sup>2</sup> “David Chalmers is the captain of the B team, along with Nagel, Searle, Fodor, Levine...and many others). He insists that he just *knows* that the A team leaves out consciousness. It doesn't address what Chalmers calls the Hard Problem. How does he know? He says he just does. He has a gut intuition, something he sometimes calls a ‘direct experience’” (Dennett, “The Fantasy of First-Person Science”, unpublished).

<sup>3</sup> *TCM*, p.xii.

The most important arguments in *TCM* appeal to logical supervenience as the benchmark for successfully explaining consciousness. A phenomenon supervenes on the physical if (and only if) any change in its properties entails a change in the arrangement of matter by which it is instantiated, while changes to or differences between substrates underlying such a phenomenon need not entail a change in how it ought to be described. Supervenient properties are higher-level than those properties characterising the physical world most fundamentally, but comprise the majority of *facts* with which we have any acquaintance, the world of “moderate-sized dry goods”<sup>4</sup>. So there is deference here to fundamental physics as the discipline revealing what is really going on in the universe, cataloguing entities whose properties are irreducible to anything more fundamental, whereas the supervenient properties only exist because the lower levels allow them to.

It was deference of this kind which seems to have motivated the mid-20<sup>th</sup> Century shift towards physicalist philosophies of mind, and delineated a significant task of the modern metaphysician to be showing how our common-sense concepts might mesh harmoniously with the image of nature revealed by physics.<sup>5</sup> Philosophy was then thought to help pave the way for a logically coherent, and ultimately intuitively satisfying, worldview. While it was not thought that the underlying causes of a higher-level natural phenomena need be knowable *a priori*, the hope was that a completed scientific picture of things would allow us to ‘read off’ facts about, say, the behaviour of lightning, given all the underlying facts regarding the trajectories of electrons, atoms and other fundamentals. Natural phenomena explicable in this fashion may be said to supervene logically.<sup>6</sup>

---

<sup>4</sup> See Austin (1962): “...does the ordinary man believe that what he perceives is (always) something like the furniture, or like these other ‘familiar objects’ – moderate-sized specimens of dry goods?” (*Sense and Sensibilia*, p8).

<sup>5</sup> See, e.g. Wilfrid Sellars’ (1963) aim to construct a “stereoscopic” viewpoint encompassing both “manifest” (common-sense) and scientific “images.” (“Philosophy and the Scientific Image of Man”, p5) Sellars identified philosophy as the attempt to show how things “hang together in the broadest possible sense” (*ibid*, 1).

<sup>6</sup> *TCM* 40-42.



Central to scientific representation is its capacity to “demystify” entities<sup>7</sup> via “extrinsic” forms of explanation, showing their behaviour to be reducible to “structure and dynamics,”<sup>8</sup> meaning how they are put together and how their components interact. For instance:

All it means for an organism to learn, roughly, is for its behavioural capacities to adapt appropriately in response to environmental stimulation. If we explain how the organism is able to perform the relevant functions, then we have explained learning (*TCM*, p44).

This simple conception of the universe as matter in motion, mysterious only by virtue of the quantities of material involved, and the consequently vast network of causal relations it instantiates, has supplanted much of what could be termed qualitative explanations of events, in terms of their meaning, purpose, or vital spirit; the kind of explanations favoured once upon a time by metaphysicians in the Middle Ages. Science’s claims to authority today are such that the possibility of extrinsic explanations for any and every event can by and large be taken for granted, with the actual achievement of such a grand project (a ‘theory of everything’, as it has come to be known<sup>9</sup>) consigned to matters of detail.

But in order to make way for mind-brain supervenience philosophers had to show that we have no first-person acquaintance with *what* the mind is – only what it does. In keeping with Chalmers’ comments above on learning, this project started with behaviourist psychology, which identified the mind itself with the disposition to behave in the right ways,

---

<sup>7</sup> *Ibid*, p42, p49.

<sup>8</sup> *Ibid*, p35.

<sup>9</sup> “...string theory is sometimes described as possibly being the ‘theory of everything’ (T.O.E.) or the “ultimate” or “final” theory. These grandiose descriptive terms are meant to signify the deepest possible theory of physics – a theory that underlies all others, one that does not require or even allow for a deeper explanatory base...[i]f you understand everything about the ingredients, the reductionist argues, you understand everything” (Brian Greene (2000), *The Elegant Universe*, p16). In a more philosophical vein, Greene continues: “The ultimate theory would provide an unshakeable *pillar of coherence* forever assuring us that the universe is a *comprehensible* place.” (*ibid* p17, emphasis added)

but study of the brain increasingly saw mental states identified directly with brain states, without any reference to behaviour.<sup>10</sup> Chalmers regards all such variations of physicalism as indefensible, however. In the wake of influential papers by the aforementioned 'B-team' of philosophers<sup>11</sup>, it has begun to look as if we know what it is we are talking about when it comes to consciousness, and that whatever it is cannot be physical<sup>12</sup>. Chalmers' contribution to the debate, as he sees it, is to systematise his predecessors' misgivings about physicalism into an argument against the logical supervenience of the mental upon the physical. He therefore challenges the idea that physics alone can give us a theory of everything. He claims that since the existence of the phenomenal character or *qualia* of experience is in no way logically entailed by the physical facts constitutive of biology, chemistry and physics, whatever it is that brings qualia into existence must be beyond the purview of current science, being solely the remit of metaphysicians to make sense of.

Nevertheless, in neither Chalmers' case nor that of most of his physicalist opponents is the possibility of an intuitively satisfying 'God's eye view' on things ever in doubt: Chalmers' position just gives metaphysicians a slight edge over physicists in attaining it. For instance, the mind-independent world is thought by Chalmers to be more or less as contemporary science represents it, with or without the amendments he proposes.<sup>13</sup>

This much about the scientific worldview seems *a priori*: it presupposes a 'picture' of how things are, as mechanisms able to be taken apart, reconfigured, and predicted, or else

---

<sup>10</sup> See the canonical defence of the mind-brain identity theory of mind in Smart (1959). Smart analyses experience in terms which remain "topic-neutral" as to *what* experiences are in and of themselves: "I...elucidate colours as powers...to evoke certain sorts of discriminatory responses in human beings. They are also, of course, powers to cause sensations in human beings...these sensations, I am arguing, are identifiable with brain processes" (p149).

<sup>11</sup> See esp. *TCM* chapters 2 and 3.

<sup>12</sup> This holds true, argues Chalmers, even if what we are talking about cannot be expressed: "We have no independent language for describing phenomenal qualities...there is something ineffable about them. Although greenness is a distinct sort of sensation with a rich intrinsic character, there is very little that one can say about it other than that it is green." (*TCM* p22)

<sup>13</sup> This will prove to be a somewhat tendentious point on Chalmers' part, since it depends on him accepting epiphenomenalism with regards to these experiential 'amendments' to the physicalist worldview; see below.

assigned to the physicist's table of irreducible elements. Thus, while emphatically engaging with sophisticated materialist accounts of consciousness, Chalmers argues for their failure on similarly *a priori* grounds: we can imagine having a completed theory of everything, and even complete knowledge of each and every subatomic particle (*TCM* 76), and still lack empirical, verifiable evidence for what is right in front of us, our subjective, perhaps ineffable sense of what experience is like. We can know *a priori* that the extrinsic or dispositional explanations science furnishes do not tell the whole story, that science only gets us as far as representing a *zombie world*, "a world physically identical to ours," down to the last particle, "but in which there are no conscious experiences at all" (*TCM* p94), and that God would have had more work to do, so to speak, after making such a universe before it contained conscious experience (*ibid* p38).

Given the reasonable assumption that his materialist opponents are conscious<sup>14</sup>, Chalmers accuses them of ignoring the obvious, or of explaining it only by virtue of a bait-and-switch between phenomenal and functional definitions of consciousness (*ibid*, p165). He levels this accusation not on the basis that their theories of the mind are empirically mistaken or even incoherent, but because he has introspected and found them to be false. The whole debate between A and B teams, essentially one of how far scientific authority extends (can it correct me if I think I am in pain?) thus amounts to a clash of intuitions: argument is to some extent secondary. Chalmers is aware that this amounts to a stalemate<sup>15</sup>, though *TCM* nevertheless offers guidance on how physicalist philosophers might rationally engage him, by showing zombies to be logically impossible.<sup>16</sup>

---

<sup>14</sup> Though he does ask; "[w]hat might be going on when [Dennett] claims that introspection reveals only judgments? [i.e., false judgements as to the existence of qualia] Perhaps Dennett is a zombie". *TCM*, p190. John Searle (1997) jokes along similar lines in *The Mystery of Consciousness* during his reply to Daniel Dennett, p120-30.

<sup>15</sup> "After a point, it is difficult to argue across this divide, and discussions are often reduced to table pounding". *TCM* p. xiii.

<sup>16</sup> *TCM* 37-8.

## **[1.2] Ghosts versus Zombies**

Framing the mind-body problem by appeal to zombies has recently been criticised by Philip Goff, a philosopher otherwise highly sympathetic to Chalmers' overall project. Goff is of interest in no less than four respects. Firstly, he tries to fully develop the implications of Chalmers' sketch of an alternative to the prevailing physicalist orthodoxy (see [1.3]). Secondly, he finds aspects of Chalmers' theses to undermine the aim of defending phenomenal consciousness against sceptical critics, who are prepared to describe its apparent non-physical properties as an illusion of some kind. Thirdly, Goff then finds himself defending claims of an increasingly far-fetched nature which harbour problems of their own, and which, additionally, may not ultimately fare any better than Chalmers' metaphysics in upholding the reality of what we ordinarily think we mean by conscious experience. Finally, Goff is more thoroughly Cartesian than Chalmers, and frequently draws on Descartes' methods for arriving at the claims he does, which turns out to be highly significant. The two philosophers are otherwise in agreement, and will be compared and contrasted during the following exposition.

### **[1.2.1] The Possibility of Having 'All the Facts'**

20<sup>th</sup> Century analytic philosophy of language, descended from Bertrand Russell and Gottlob Frege, gives a number of reasons to be wary of the kind of metaphysics Chalmers is doing. The first comes from Frege himself, whose seminal paper "On Sense and Reference" (1948) addressed how names relate to objects. In response to the problem of how the names 'morning star' and 'evening star' could both turn out to refer to the planet Venus, Frege

distinguished between the sense of a name and its reference, whereby the former picks out a mode of presentation or way of thinking about the latter. Although Frege construed sense as something akin to Platonism's concept of mathematical facts as mind-independent abstract objects, reference itself was made possible by the mental act of grasping a sense. Assuming all this, a physicalist might wish to argue that phenomenal appearances are merely the sense by which we subjectively achieve reference to objective brain processes, and treat mind-body dualism as being on a par with the erroneous dualism of morning and evening stars. But Chalmers denies that qualia have a sense-reference distinction; the referent *is* its mode of presentation, and therefore can be exhaustively known a priori.

Later arguments from W.V.O. Quine sought to dispute the a priori/a posteriori distinction altogether, but I will set these aside, since they were challenged in turn by Kripke's influential *Naming and Necessity*. Kripke rejected "the conception of reference... as dependent on, and determined solely by, the individual language user's knowledge, intentions and dispositions" (Noonan 2013, p1), for instance the descriptive content with which they associate an object. Rather than identifying names with definite descriptions (as he interpreted Frege and Russell to have done), Kripke views them as rigid designators whose reference is completely independent of what, if anything, a speaker believes about them. Reference is instead achieved when it involves a genuine causal connection between the speaker and the object in question (see Kripke 1970, p91), which might only be discovered a posteriori. The identity of mind and brain – a necessary truth, if true at all – may therefore be an a posteriori necessity.

Kripke nevertheless finished *Naming and Necessity* with an argument against a posteriori physicalism: since the essence of qualia is their appearance, their physicality must be apparent a priori. Since this is not so, 'pain' cannot refer to 'c-fibres firing'. According to Chalmers, this argument failed, because Kripke only established the logical possibility of

disembodied pains, without showing that that possible world also happens to be our own. The crucial point is that “what is essential to that pain state is its feel, and only its feel. But such claims about the essential properties of individuals are hard to justify” (*TCM* p148). Consequently, Kripke’s argument could be read against his wishes as deepening the physicalist case after all. All evidence indicates we do not live in a world in which disembodied mental states happen to be possibilities; though the types of entities we refer to as pain could have been non-physically instantiated, there is no example of pain tokens in the absence of corresponding physical states.

In response, Chalmers proceeds with a discussion and defence of the distinction between primary and secondary intensions, following Roderick Chisholm’s analysis of intentionality in terms of ‘intensional’ sentences whose truth conditions are not necessarily preserved by the substitution of co-referring terms, and which do not license existential generalisations (Chisholm 1957). As is typical of his approach to epistemological and metaphysical questions, Chalmers draws this distinction along the lines of possible world semantics, whereby primary intensions identify their referents in *this* world, while secondary intensions have the same referent across all *possible* worlds. Showing that these two types of intensions can come apart motivated the adoption of semantic externalism by analytic philosophers, thanks to Kripke and in part due to a famous thought experiment outlined by Hilary Putnam (1975). Putnam posed the question of whether we would be inclined to say that beings just like us, on some distant ‘Twin Earth’, were really referring to water on their planet if it turned out to be composed of XYZ, rather than H<sub>2</sub>O. It seems more reasonable to suppose that they mean a quite different kind of stuff by their term ‘water’ from that to which we refer. As was the case with Kripke, this conclusion might then be viewed by physicalists as undermining the philosophical importance of first-person certainty

as to what one means when referring to (e.g.) water, given that its mind-independent properties – those to which we ‘in fact’ refer – determine what we are really thinking about.

This certainly worried Chalmers, and he dedicated the beginning of *TCM* to getting the problem out of the way; he also returned to the issue more recently in “The Two-Dimensional Argument against Materialism” (Chalmers 2010; henceforth *2D*). In light of semantic externalism, the challenge for the anti-materialist is to arrive at a thought whose reference *cannot turn out to be* metaphysically benign upon further inspection.<sup>17</sup> Chalmers arrives at the following: whereas being able to imagine *pain* lacking a physical substrate only evidences the token separation of mind and matter, our imagining a complete *zombie world* enables us to rule out the logical supervenience of mind from matter altogether, in this world or any other.

The physicalist will reply that Chalmers’ appeal to a God’s-eye-view on the universe is just as dubious. There is no guarantee that one can imagine having *all* the relevant facts about neuroscience without one’s imagination being subject to illusions. Goff concedes that sceptics will deny that they can imagine zombies, and will ask philosophers to prove that they have *really* imagined what they thought they had, rather than dogmatically insisting upon it (He calls this the “wait and see response”<sup>18</sup>; Chalmers (*TCM* 162) calls it “don’t-have-a-clue materialism”). While not especially constructive, responding in this fashion might seem appropriate given the state of modern biology; Patricia Churchland warns against “inventing an explanatory chasm [between mind and matter] where there really exists just a broad field of ignorance” (she takes the fact that most people can sign their name with their

---

<sup>17</sup> Goff is likewise suspicious of any concession, however small, to the idea of there being an a posteriori identity between consciousness and the material world, and charges Chalmers with failing to fully appreciate his own complicity with this line of thinking. See below.

<sup>18</sup> Goff (2012), p742

feet in a style that is recognisably theirs as a 'hard problem' for current neuroscience, for instance).<sup>19</sup>

In light of the above, the materialist critic of zombies might be better off treated as a kind of sceptic, and the response to them should proceed via reflection on the limits of scepticism, rather than on the nature of possibility and its links to conceivability.<sup>20</sup> Inspired by Descartes, Goff will then go on and draw ontological conclusions on the basis of his response to the sceptic's challenge. Descartes famously thought he had demonstrated that the mind is first and foremost a "thing that thinks", whose conceivability is a uniquely reliable guide to its necessary existence at the moment of doubt (since doubt is a type of thought).<sup>21</sup> In Goff's hands, such an entity, which he calls a "ghost", is defined as "a thing such that there is something that it is like to be that thing"<sup>22</sup>, focusing the issue of the hard problem onto the existence of qualia; which somewhat distances him from Descartes' intentions, although not from the way Descartes has been received.<sup>23</sup> More recently, Goff has characterised ghosts as "o-phenomenal" or "o-experiential" facts, or "the phenomenal properties we pre-theoretically associate with humans and other animals" (the 'o' stands for 'ordinary', but is also sometimes suggested to be 'organism'; see below).

---

<sup>19</sup> Churchland (1996), p405.

<sup>20</sup> See *TCM* 195-6 for a similar characterisation of the physicalist as a sceptic; one who doubts the reality of consciousness. Chalmers thinks that the kind of scepticism one would need to entertain before seriously considering that we might be zombies is too extreme to be taken seriously, and does not treat his physicalist opponents as being generally committed to such scepticism. His focus is on philosophers who think realism about consciousness is compatible with physicalism, i.e., those who downplay whether zombies are conceivable or whether their conceivability is relevant. However, the more extreme, sceptical line has enjoyed resurgence recently; I say more about this in chapters 2 and 5.

<sup>21</sup> Descartes then goes on to infer the necessary existence of God from his sheer conceivability, as well – but the 'discovery' of the I that thinks is the more fundamental of the two.

<sup>22</sup> Goff (2010), p124.

<sup>23</sup> Descartes (1984) writes: "...as for movement and sensation, I refer them to the body for the most part, and attribute nothing belonging to them to the soul, apart from the element of thought alone" (Fifth Set of Replies: §351, p.243). Daniel Hutto interprets this to mean that "Descartes was not concerned with consciousness...he would not have recognised its qualitative character as being the motivating factor behind his dualism. His focus was on the irreducibility of the intellect, not on the character of experience." (Hutto, 2000, p3). See also Cottingham (2000).



Whereas Chalmers used the thought experiment of a completed physics to conclude science could never prove whether we are living in a zombie world, Goff is more circumspect:

Defining conceivability this way is not entirely uncontroversial. Chalmers [2002b] calls it 'negative conceivability', and contrasts it with 'positive conceivability', which is defined in terms of what can be positively imagined. However, the notion of positive conceivability is problematic for a number of reasons. Firstly, the relevant notion of 'imagination' is obscure. We are not simply talking about sensory imagination, as this would make too many states of affairs inconceivable, e.g. space being infinite, or there being four-dimensional objects. But nor is it simply thinking a thought involving the proposition, as this would make too many states of affairs conceivable, e.g. the state of affairs of there being a square circle. It is difficult to find a middle way between these two extremes. Secondly, there is certain dialectical advantage in setting things up in terms of negative, rather than positive, conceivability. It is difficult to have an argument with someone over whether or not a certain state of affairs can be imagined. I say I can imagine a zombie, you say you can't. It's difficult to know where we go from there. It is much easier to have an argument over whether a state of affairs is contradictory or incoherent. (Goff 2009b, p295)

Goff poses his thought experiment as negatively conceivable, as we cannot rule out being in a sceptical scenario, whereas we *can* rule out being wrong about consciousness. This is stronger than the zombie argument because we cannot rule out being wrong about the possibility of zombies, and cannot rule out having failed to imagine all physical facts. As I will argue in chapter 4, Goff is not always this careful to distinguish positive and negative conceivability, and the different types of imagination. Nevertheless, his early remarks quoted here provide a good methodological foundation for thinking about the hard problem,

as they underline the possibility of formulating it without appeal to Chalmers' two-dimensional 'modal rationalism'<sup>24</sup>.

### [1.2.2] Varieties of A Priori Metaphysics

Chalmers and Goff are both driving at the same conclusion, that whatever explains the existence of consciousness is not going to be a further *empirical* fact, irrespective of whether such facts could be arrived at via armchair reflection. They diverge on the issue of how to motivate it, with Chalmers having successfully put zombies into the spotlight, and Goff favouring ghosts. The zombie argument invites us to pretend we have "all the facts" about the universe before us; not something easily imagined. On the other hand, everybody with a cursory knowledge of philosophy should be able to grasp Descartes' train of thought in the *Meditations*. Unlike Descartes, however, Goff does not take it to be the special task of philosophy to *justify* knowledge using only what is available to a single, thinking subject, and so avoids the Cartesian project of having to single-handedly prove the existence of God so as to ensure that each individual's efforts to know the world are not in vain. Instead, he makes the more modest claim that common sense, empiricism and appeal to theoretical virtues such as parsimony and cohesion with existing theories are needed in order to understand reality, but that the datum of o-consciousness provides an "unrevisable" "fixed point" of reference for doing so (Goff 2014b, p4). He agrees with Chalmers that semantic externalist considerations give us good reason for thinking that solipsism is not in need of refutation before knowledge of the world at large is possible. What we think or mean, in referring to things in the world, can only be understood by first presupposing that the world exists

---

<sup>24</sup> Chalmers identifies himself as a modal rationalist, one who thinks the space of possibilities is not in any way foreclosed to maximal (ideal) rational reflection, in *2D* p184. This need not imply that we possess the capacity for ideal reflection about any issue ourselves, of course.

independently of ourselves.<sup>25</sup> Working inwards, empirically deriving those few facts known to a subject in the grip of solipsistic doubt (e.g., the phenomenal or qualitative component of the subject's *narrow content*), is the real problem.<sup>26</sup>

It is significant that Chalmers, who writes that "I am [...] giving myself the physical world for free" (*TCM* 75) finds objections to the possibility of genuine machine intelligence unpersuasive<sup>27</sup> – as suggested by his previously quoted remarks on what it means to learn something. If meaning is mind-independent, then the possibility of 'strong' AI follows automatically; an intelligence, artificial or otherwise, would simply have to be in the right kind of relationship with the external world to count as a knower, irrespective of whether it knows that it knows anything (it could even be a convinced solipsist). Because the presence or absence of consciousness is, for Chalmers, a totally separate issue from the presence or absence of the kind of functional architecture supportive of intelligence, he is not threatened by John Searle's 'Chinese Room' argument (1980) against likening the mind to a computer, and can consequently steer himself clear of disputes with cognitive/neuroscientists for whom this supposed likeness has been the basis of their research.<sup>28</sup> Against Searle's protestations, even a computer implemented using punch-cards and a tireless slave could, Chalmers argues, be said to understand Chinese, so long as it could reliably issue what we would regard as appropriate responses to conversation partners fluent in the language. The cognitive phenomenology – what it is like to understand – would, so to speak, 'float on top' of the Chinese Room, though as will be seen in the next section it would be misleading to put it in such dualistic terms.

---

<sup>25</sup> See e.g. Hilary Putnam's argument against neo-Cartesian 'brain in a vat' scenarios in Putnam (1981), pp. 12-17.

<sup>26</sup> See also the latter half of Mark Rowlands' *Externalism* (2003) which addresses the challenge of narrow content to externalism.

<sup>27</sup> *TCM* n.10, p360.

<sup>28</sup> As Bechtel (2008) argues. As will be seen, Goff, is much less careful to avoid treading on scientists' toes.

Chalmers concedes that there is something strange about this – the “paradox of phenomenal judgement”<sup>29</sup>. In effect, he thinks that there are two selves corresponding to the two descriptions of what exists – a functional self described by psychology and neuroscience, and a phenomenal self which is intuitively self-evident. One makes claims about being conscious, but might (in some possible world) be a zombie; the other knows consciousness directly. The *existence* of the phenomenal self coincides with its self-knowledge, but as such this kind of knowledge differs from that involving the deployment of concepts, or the possibility of expressing oneself, *including to oneself*.<sup>30</sup> By contrast, identifying the functional self that judges itself conscious with the phenomenal self clears the way for physicalism. Chalmers discusses this by reference to an introspectively-limited AI, one that is duped by its programming into believing the operating system’s symbols by which it accesses its own states possess intrinsic properties, because it has no need to represent itself in much detail.<sup>31</sup> Chalmers states that he finds the possibility to be the closest to a knock-down argument for physicalism, because we cannot prove we are not in the same epistemic situation ourselves: the apparent irreducibility of qualia could be an artefact of our ‘operating system’. Chalmers replies that this type of psychological explanation doesn't allow us to deduce the existence of qualia; we can imagine a zombie making the same mis-judgement as to the nature of its (non-existent) experiences.<sup>32</sup> So long as his interlocutors are not thoroughgoing sceptics, and agree that the first-person data is in need of explanation rather than proof, Chalmers does not seem to be in danger.<sup>33</sup> And his distinction between the functional and phenomenal self allows him to concede the force of

---

<sup>29</sup> TCM 177

<sup>30</sup> This marks a break from Kripke’s theory of meaning: though they are rigid designators, phenomenal concepts are not causally mediated.

<sup>31</sup> *Ibid* 184-8. See also Dennett, *Consciousness Explained* (1991), p311-2, which defends the ‘user illusion’ explanation of consciousness which Chalmers rejects.

<sup>32</sup> See Chalmers (2010), p312-20.

<sup>33</sup> Physicalists who refuse the burden of deducing qualia from the physical facts are discussed in chapter 4. In any case, the paradox of phenomenal judgement casts a shadow over anti-physicalism, and I will return to it again in chapter 2.

the argument (which, really, can be traced at least as far back as Freud) that humans and deluded AIs might have a lot in common so far as knowing our own psychology goes.

In pushing the semantic externalist line, Chalmers even downplays the significance of *subjectivity*, in the sense of having a 'point of view' on the world, holding it to be a trivial matter of where one happens to be physically located. After all, having that fact to hand "will not enable [a colour-blind colour scientist] to know what it is like to see red," (*TCM* 133; see also 144) which is the real nub of the hard problem. As such, he construes qualitative experience as a strangely objective feature of the universe, one which, for all we know, is capable of existing in the absence of a body, conceptual thought, or intelligence of any kind.<sup>34</sup> Chalmers can therefore be compared with philosophers of the past who denied the existence of a subject lurking behind, or perceiver of, subjective qualities. See e.g. Hume (1960), who confesses to find in introspection merely "a bundle or collection of different perceptions" (251), but no bearer of them. And in the 19<sup>th</sup> Century G.C. Lichtenberg put a grammatical gloss on Hume's position; "It thinks, we should say, just as one says, it lightnings" (Lichtenberg 2012, p152). It will be argued that Goff fails to fully shake off this particular picture of the mind, in spite of prioritising subjectivity via his appeals to Cartesian self-certainty. In fact, being able to picture the mind at all is part of Goff and Chalmers' shared conception of consciousness; they allow for the possibility of a God's-eye-view in which all experience, across time and space, could be viewed at once in the manner of a television network seen from inside a master control room; a clear and distinct conception of all that there is, physical and mental alike.

Chalmers and Goff's externalism and fallibilist appeal to natural science is heavily offset by the role a priori theorising continues to play for them. This would be all but

---

<sup>34</sup> For example, Chalmers (*TCM*, p353) rejects the intuitive association between personal identity and consciousness, and concludes that "each of tomorrow's minds are equal candidates to count as me, and there is no fact to distinguish them".

guaranteed by their anti-physicalist metaphysics, but their approach to philosophy is indebted to the physicalist David Lewis, whose deployment of the so-called “Canberra plan” is particularly evident in Goff’s metaphysics (see Goff 2010, p130-131). The plan is named after the fact that Australia’s capital of Canberra was constructed midway between Sydney and Melbourne as a compromise for the two cities’ prior inability to decide which of them should be nominated as capital. It is a city built by and for government bureaucrats, and its association with a certain style of philosophy, characterised by ad-hoc compromise solutions to difficult problems, was originally intended to be pejorative (Braddon-Mitchell and Nola, 2009, p1). By the time it was coined, however, the Canberra plan had been employed for decades, by (among others) David Lewis and Frank Jackson, for whom the aim was primarily in reaching a compromise between the appeal of common sense and that of physicalism. “The first step will be to collect together the “platitudes” concerning the *X* to be analyzed... If there is a sufficiently unified set of agreed platitudes about *X*, then there will emerge a theoretical role for the central notions describing the domain in which we are interested.... The second step ... will be to discover what in the world, if anything, plays the roles so described” (*ibid*, p7). Anti-physicalists broadly sympathetic with this plan will, of course, argue that satisfying the theoretical role conscious experience plays in our worldview must take us beyond physicalism.

How exactly such an analysis is to be carried out will depend to a large degree on appeals to Ockham’s razor. As Daniel Nolan (2005) sees it, for his part “Lewis seems to be engaged in the project of accounting for as much as possible with as few resources as possible” (15-6). Potential candidates for cashing out notions such as experience, free will, beauty, causation, etc. will be assessed on the basis of their elegance, simplicity, and explanatory scope (Braddon-Mitchell and Nola 2009, p9). And this in turn leads Lewis to favour simple ontologies, with physicalism in effect falling out of his conviction that natural

kinds, having been rehabilitated by Kripke, exist only at a microscopic physical level: the supervenience relation, which Lewis helped popularise, treats higher level properties as simply abbreviations of relations between the natural properties. Without such a relatively sparse array of natural properties, Lewis argues, there could be no particular content to propositions such as “everything is x”, or the possibility of meaningfully abbreviating particular events that take place, via lawlike generalisations (Nolan 2005, p84-6); nor could the common-sense platitude that some things resemble each other more than others be cashed out (*ibid* p22-3; Lewis 1983).

### [1.3] Russellian Monism

In the later stages of *The Conscious Mind*, David Chalmers tentatively defended a form of panpsychism, the claim that consciousness is to be found indiscriminately throughout the animate and inanimate universe. Made famous in analytic philosophy by Bertrand Russell – hence ‘Russellian’ monism,<sup>35</sup> this doctrine has variously been identified, over the years, with neutral monism (when the nature of reality is argued to be *neither* mental nor physical), panprotopsychism (where the nature of reality is a primitive or pseudo-consciousness upon which our experience supervenes) and finally, in Goff’s terminology, “funny physicalism”.<sup>36</sup> Chalmers describes his position as fitting “the letter of materialism”, while sharing “the spirit of anti-materialism”.<sup>37</sup> Goff glosses the central claim as a response to the inadequacy of

---

<sup>35</sup> Russell (1927). However, variations of it can be found throughout philosophical and religious history: see David Skrbina’s introduction to Skrbina (ed.), *Mind that Abides* (2009), pp1-32. The doctrine’s modern resurgence may be more accurately traced to William James, who has sometimes been characterised as a neutral monist or panpsychist, rather than to Russell; see Cooper (1990).

<sup>36</sup> A slightly derogatory term for what he finds to be objectionable versions of the thesis (see Goff 2010, p119); in a more recent paper (Goff 2015a) he terms it “constitutive Russellian monism”, in both cases attributing this variant to Chalmers.

<sup>37</sup> Chalmers (200a), p265..

physicalism construed as identifying what there is with “structure and dynamics” alone: “dispositional properties cannot exhaust the nature of matter...there must be some categorical nature to matter which grounds its dispositions... which is hidden from the perspective of physics”<sup>38</sup>. This is the “empirically indiscernible” realm of “physical ultimates”, and “according to panpsychism there is something that it is like to be a physical ultimate”<sup>39</sup>. Whereas empirical entities<sup>40</sup> are identified in terms of their relations to each other, their categorical nature is intrinsic and self-sufficient without needing to enter into relations at all.<sup>41</sup>

Chalmers is quite clear in *TCM* that the Russellian solution is purely stipulative<sup>42</sup>, and defends it “in the spirit of getting ideas onto the table” (*TCM* 277): he takes it to be a coherent response to the problem of fitting mind and matter back together, and that alternative metaphysics of mind – idealism, for instance – may fit the evidence and even turn out to be preferable, though they are not canvassed in *TCM*. But Russellian monism allows Chalmers to make a somewhat plausible case for *property* dualism while remaining a realist about the external world and scientific postulates such as electrons, and without appealing to violations of known physical law; scientific authority can still be safely conceded to on matters of empirical fact, in particular the claims of modern neuroscience. It allows him to explain why zombies or zombie universes are not really possible, as most philosophers believe, while using physicalism’s inability to rule them out as a *reductio ad*

---

<sup>38</sup> Goff (2010), p120.

<sup>39</sup> Goff (2009b), p289.

<sup>40</sup> Goff means fundamental physical entities, which arguably are not empirically perceivable so much as deduced from available empirical evidence. Nevertheless, the postulation of ‘ultimate’ or intrinsic properties is not treated as a falsifiable hypothesis, and the properties in question inhabit a more fundamental realm of existence than that discussed by modern physicists in arguments over what kinds of fundamental particles populate the universe.

<sup>41</sup> This definition, which all but guarantees the epiphenomenalism of categorical properties, will be challenged in the next chapter and in subsequent discussion of John Heil, who defends a form of neutral monism without nonrelational properties, and which is not property dualist.

<sup>42</sup> See p277, 305



*absurdum* against physicalist metaphysics, and (as already discussed) to give Searle's Chinese Room argument its due while denying its conclusion:

*Programs* are abstract computational objects and are purely syntactic. Certainly, no mere program is a candidate for possession of a mind. *Implementations of programs*, on the other hand, are concrete systems with causal dynamics, and are not purely syntactic. An implementation has causal heft in the real world, and it is in virtue of this causal heft that consciousness and intentionality arise. (TCM 327)

Phenomenal properties or qualia are responsible, on this view, for implementing - literally *realizing* or 'making real' - possible states of matter; any states which are conceivable (logically possible), or which do not conflict with the laws of fundamental physics (naturally possible).<sup>43</sup> Without granting any extra causal powers, the fact that something really, rather than potentially, exists, is clearly essential to it having any causal powers in the first place. Similarly, though conscious experience cannot be the cause of thought or behaviour, these being determined by the extrinsic or relational properties described by biology, it is nevertheless distinctively essential to them.

Like Plato's forms, the existence of consciousness is for Chalmers the *formal* cause of phenomenal judgements being true. Only myself and (and God, if there is one) could know if my beliefs about, and expressions of, my experience are accurate; my *stated* beliefs about red could be systematically false due to my suffering an inverted spectrum of colour-qualia or suchlike. That my beliefs' semantic content is ultimately determined by facts which would only be subject to a second opinion from an *omniscient* spectator is just another consequence of Chalmers' combination of semantic externalism and Russellian monism.

---

<sup>43</sup> On the distinction between logical and natural possibilities, see TCM 34-8.

Since talking about consciousness is determined by extrinsic, causal relations, rather than intrinsic phenomenal properties, embracing the paradox of phenomenal judgement commits Chalmers to epiphenomenalism. However, his panpsychist metaphysics allows him to stipulate that a disconnection between one's private and publicly verifiable beliefs about what experiential states one is in would never happen, due to an underlying symmetry between states of consciousness and information states in the brain and nervous system.<sup>44</sup> This in turn commits him to thinking of phenomenal experiences as somehow constructed out of whatever it is which enters into informational states – some sort of basic building block underlying all possible experience. Goff, however, is critical of such reductionism, treating it as a major concession to physicalism that runs against Russellian monism's revolutionary potential for the philosophy of mind.

Goff proposes that we get a good grasp of Russell's causally isolated realm of matter 'from the inside', by following Cartesian doubt through to its conclusion, finding that one cannot doubt having experiences as of whatever is seemingly present to oneself. At that point we arrive at a conception of something whose existence is exhausted by its experiential properties, a pure sense of one's own being without any knowledge of what it is whose existence is being 'realized'. Russellians treat this introspective evidence as a sort of inverted topic-neutral description of the brain and nervous system, which he takes to comprise the 'what' realized by those conscious states with which we are acquainted.<sup>45</sup> By contrast, the physicalist J.J.C. Smart took conscious states to be the 'what' realized by neurons<sup>46</sup>. As Goff puts it, "In its most elegant form Russellian monism is a kind of

---

<sup>44</sup> *TCM* 275, 285-6. Chalmers concedes on p308 that more work needs to be done to convincingly rule out inverted spectra, via careful examination of the kinds of laws that could possibly link phenomenal and physical information states.

<sup>45</sup> As does Galen Strawson, who similarly sees Cartesian scepticism as our 'way in' to transparently conceiving the brain; see his "Real Materialism" (2008), p21.

<sup>46</sup> Michael Lockwood also links Russellian monism to an inversion of Smart's thesis in his *Mind, Brain, and the Quantum*, (1990) p160.

unorthodox identity theory... pain is identical with c-fibres firing, but the real nature of the state is understood only when it is thought of in phenomenal terms” (Goff 2015a, p373-4).

But Goff is in fact opposed to this ‘elegant’ picture, and views Chalmers’ metaphysics as the last bastion for philosophers keen to keep off scientists’ home turf:<sup>47</sup> hence the label of “funny physicalism” to describe Chalmers’ position. This section will consider his criticisms made in papers between 2006 and 2010, when he was most under the influence of Descartes and, he claims, Aristotle.<sup>48</sup> At this time period he was especially polemical, arguing that the hard problem needs to be more than an amendment to the modern scientific worldview, before it can overcome the problems which Chalmers is certainly aware that it faces<sup>49</sup>: the vast number of qualia associated with non-living entities, the contingency of mind-brain relations which have to be stipulated away, and the associated problem of individuating a subject that would circumscribe a limited set of conscious experiences in amongst the anonymous buzz of qualia making up the universe (a variant on the Sorites paradox). These are not problems unique to Chalmers’ particular variation of panpsychism, which is deliberately accommodating to reductionist physicalism due to regarding the brain as logically supervenient on particle physics. Similar complaints go back to William James, whose criticisms of panpsychism are regarded by Goff as well as contributors throughout Skrbina’s (2009) edited volume on the topic as the most serious challenge to its coherence:

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

---

<sup>47</sup> Goff (2015a), p388-9

<sup>48</sup> Fundamental confusions in his position around this time will be examined in [1.4], with these likely being responsible for his change of heart.

<sup>49</sup> See *TCM*, p300-310

consciousness belonging to the group as such should emerge. And this 101<sup>st</sup> feeling would be a totally new fact; the 100 feelings might, by a curious physical law, be a signal for its creation, when they came together; but they would have no substantial identity with it, not it with them, and one could never deduce the one from the others, nor (in any intelligible sense) say that they evolved it. (James 1983, p162)

The hope expressed in *TCM* is that panpsychism raises questions of a sufficiently obscure nature that philosophers would be wise to keep their options open. Goff is for the most part more pessimistic, and his reasons for this seem to follow from his close adherence to Lewisian metaphysics, whereby the world consists fundamentally of point-sized natural properties, a “mosaic” of “local, particular matters of fact” (Lewis 1986, p.ix) with spatiotemporal relations to one another.<sup>50</sup> Goff points out that from the first person, I do not find a cloud of simple, loosely associated qualia waiting for me at the end of Cartesian doubt. And from the ‘God’s eye view’ from which panpsychism would appear true, one could not deduce evidence of any ghosts. Treating the universe as kind of paint-by-numbers kit coloured in by patches of experience (one for each particle) leads to the conceivability of a *panpsychist* ‘zombie world’ lacking ghosts altogether.<sup>51</sup> Lewis’s characterisation of the scientific conception of the universe as a cloud of particles, does not seem to form a rich enough supervenience base to explain the kind of holistic, integrated experience we are familiar with, turning the issue where panpsychism is concerned into one in which the usual range of physicalist options (reductionist, emergentist, eliminativist, etc) are re-created, this time by reference to a metaphysical or non-observable, rather than physical or neurochemical, level of reality consisting of micro-or proto-phenomenal properties.

---

<sup>50</sup> Lewis’s ontology here will be critically discussed in more detail in the next chapter.

<sup>51</sup> Goff (2009b) p296. Goff uses the imagery of mind as a kind of paint in a later defence of (non-constitutive) panpsychism; “All we get from physics is this big black and white abstract structure which we metaphysicians must somehow colour in with real categorical nature.” (2017b, p284)

Like the physicalist, then, the panpsychist has the burden of providing an analysis of first-person experience in terms of relational properties, specifically those that would make a bundle of qualia 'add up' to a whole person, when only intrinsic properties seem to be up to the task; both Goff and Chalmers having already agreed that any attempt at explaining experience in terms of something else is little more than eliminativism with better PR.<sup>52</sup> Neither of them, that is, wants to separate the reality of consciousness from what can be known about it from the first person, so there is no possibility of treating micro-to-macroconsciousness supervenience ("constitutive" or "funny" Russellian monism) as a fact one could only come to know *a posteriori*.

Nevertheless, Goff argues, Chalmers' constitutive panpsychism is exactly the sort of position liable to slide into physicalism, because it portrays macro-consciousness or *o*-properties as "less fundamental than ... the properties of fundamental physical entities" (Goff 2010, p129). Echoing Kripke's argument against physicalism, he argues that since my mind's identity with fundamental physical properties is inscrutable when conceiving of a ghost, and must be stipulated by the metaphysician, it comprises an extra, non-mental aspect of one's essence.<sup>53</sup> Having gone this far, the Russellian may then be tempted to conclude that full knowledge of *ourselves*, as much as of the intrinsic properties of all matter, is beyond us, and so be led to consider that the conceivability of panpsychist zombie worlds (those consisting of micro-, but not macro-conscious subjects) is an illusion of some kind. At which point they may as well drop panpsychism and become a full-blown physicalist.<sup>54</sup>

---

<sup>52</sup> See Goff, Goff (2015a) 389-90, Chalmers, *TCM* p165. See also Nagel's relegation of the range of reductive or non-reductive physicalist philosophies of mind to "external theories", by definition incapable of making claims about 'internal' or subjective facts. (*The View From Nowhere*, 1986, p.7)

<sup>53</sup> Goff (2010), p128-9.

<sup>54</sup> I challenge Goff's dichotomy here in chapter 6.

However, Goff's ghost argument, if successful, yields "creatures which by definition have conscious states that are not realised in any more fundamental nature" (Goff 2010, p125). Of course this assumes that a thing's essential properties are those which one cannot doubt away without changing the thing itself; since we can doubt the physical but not the mental, the mental cannot essentially have physical properties. Russell's ingenious contribution was to show how these sorts of radical doubts could be compatible with realism about mind-independent objects: rather than showing the mind to be non-physical, they show the physical to be non-empirical, because mind is stipulated to be precisely that which the physical essentially is.

In pressing the importance of o-properties, Goff is more sensitive to the gulf between the abstract subject matter treated in anti-materialist metaphysics and "the kind of conscious experience pre-theoretical common sense attributes to [people] on the basis of our everyday interactions" with them<sup>55</sup>. What is missing, he suggests, is an account of the subject's role in shaping the character of human experience, something that Chalmers, as mentioned, regards as a side-issue to be explained reductively. But despite this crucial difference, Goff is not going to abandon Russellian monism altogether, or rethink the theoretical framework in which he and Chalmers make their case. This leads to a clash between his metaphysical commitments and his Cartesian epistemology with regards to what can be known about the mind, which in turn leads him to a quite deflationist notion of what a subject is: though he denies that experience can be reduced to a bundle of contingently related qualities, as it was for Hume, one could still follow Lichtenberg and treat o-properties as instances of experience occurring without a subjective bearer. Goff simply thinks of the bundle as necessarily unified in a way which clashes with the metaphysics of composition to which he (like Lewis) believes science is committed.

---

<sup>55</sup> Goff (2009b), p390.

#### [1.4] Goff in Focus: Foundations

Goff has developed a fairly consistent offshoot of Chalmers' polemic against physicalism and programmatic remarks on panpsychism over the past decade. Scratching beneath the surface, however, does reveal a few changes over that period which are not always acknowledged, and reveal inconsistencies in his position, discussed below. The trajectory of Goff's thought may be summarised as follows. His earlier papers (e.g. 2006, 2009a, 2012) tended to take a more cautious view with regards to metaphysics, and focused on bolstering the case of o-properties' irreducibility to their parts. Physicalism and constitutive panpsychism came under attack at this stage, but Goff seems to have toyed with several different ways of developing the arguments there into a positive metaphysics of his own, including taking positions more reminiscent of substance dualism. Already suggested in some comments in Goff (2010), a defence of dualism and free will is given in a 2013 draft of Goff (2015a) titled "Against Funny Physicalism", and free will (but not dualism) in "The Phenomenal Bonding Solution to the Combination Problem" (published in 2017, but drafted in 2014).

The latter paper is significant as it defends 'unrestricted phenomenal bonding', which significantly dilutes the relationship between o-properties and the body (see below). His 2014a and 2014b follow suit, but avoid or reject the issue of free will as extraneous to the metaphysics of consciousness. This is not to say that Goff was becoming less speculative, or more confident in mainstream science's ability to say anything informative about consciousness. If anything the opposite is the case, as the focus was by now heavily on the metaphysics of phenomenal bonding, which Goff maintains is beyond empirical investigation. His recent (2017) book has finally presented a complete picture of his views on

bonding, but it also to some extent recapitulates his earlier arguments for dualism and the irreducibility of mental causation. This is not quite as far-fetched as it sounds, as his solution, ‘cosmopsychism’, is able to accommodate the lack of scientific evidence for irreducible mental substances and powers without engaging in selective scepticism, as his earlier flirtations with these themes did. Goff has tended to be clear about how his views on the issues of bonding and causation have changed over time. Nevertheless, there is in addition an unacknowledged clash between his epistemology and metaphysics of consciousness, which can be understood as follows.

One constant throughout Goff’s writings is metaphysical essentialism. Building on his earlier appraisal of Lewis, Goff also appeals to Theodore Sider’s recent (2012) neo-Lewisian contributions to essentialist metaphysics, according to which a “joint-carving expression” is one employing concepts that hook onto things in themselves. Criteria for representing reality include the theoretical utility of concepts like ‘electron’ compared to, e.g. ‘grue’ in helping us describe the truth conditions for the existence of objects and states of affairs.<sup>56</sup> Simply being able to trace the causal history of a concept’s use back to some item thus ‘baptised’ as ‘elephant’ or ‘stock market’ – as in Kripke’s original variant of essentialism – falls far short of what Sider would regard as ‘joint carving’, a privilege he reserves for propositions involving only fundamental physical entities. This is because, as Goff puts it, English “is more influenced by everyday usage than the metaphysical structure of the world” (Goff 2015a, p385). According to Sider, specifying the metaphysical truth conditions for a concept such as ‘table’, would not draw on the causal history of how the word has referred, but would instead consist in a list of all the arrangements of particles by which the sentence ‘there are tables’ comes out true - using only expressions drawn directly from fundamental physics. Understanding the truth conditions underlying everyday life is therefore an ideal

---

<sup>56</sup> Goff (2015a) 384-6. Goff also frequently refers to disjunctive properties, e.g. “being and elephant or an electron” as unfit for joint-carving purposes (*ibid*, and Goff 2010, p130).



which we can never live up to, although Sider thinks we can gesture in the right direction (Sider 2012, p116–18).<sup>57</sup>

While Sider is ultimately committed to “the non-existence of composite objects”<sup>58</sup>, Goff’s main interest is in the ideals of clarity, distinctness, and theoretical utility - for metaphysicians - with which Sider motivates his reductionism. More recently, for instance, Goff (2017a) has entertained cosmopsychism, in which whole universe would be fundamental. In either case, having arrived at a transparent conception of o-properties, and having ruled out reducing consciousness to facts about non-conscious entities, we are led to the conclusion that the unity of experience reveals a joint-carving property. When Goff writes that “we can understand a ghost to be *a pure subject of experience*: a creature whose being is exhausted by its being conscious, by there being something that it is like to be it” (Goff 2010, p123), he means that alongside the essential *physical* constituents of the universe, and their own micro-phenomenal ‘intrinsic’ properties, there is another layer of *non-physical* ‘ultimates’ (he calls this “layered monism”<sup>59</sup>), and that it is *these* whose intrinsic qualitative properties are unveiled to us.

Bearers of these properties are called ‘o-subjects’, and are regarded as metaphysically basic, alongside the fundamental physical properties of “mass, charge, spatiotemporal position, properties characterizing the distribution of various spatiotemporal fields, the exertion of various forces, and the form of various waves” (TCM p33) which Chalmers of course appends merely with micro-phenomenal properties. However, ghosts or o-properties are *not* simply brains ‘from the inside’, at least if ‘brain’ refers to composite

---

<sup>57</sup> These points all recall arguments by Lewis on the nature of composition; see Lewis (1986), p212-3, and Nolan (2005), p34-6.

<sup>58</sup> Goff (2017b), pp283-304; n.7, p303

<sup>59</sup> “Against Funny Physicalism”, p8. At this point he seems to have been favourable to layered monism. Goff (2017a, p151) later identifies the view with emergentism; by which time he was critical of the view for its failure to address mental causation. Chalmers (2017, p192-3) goes even further and compares emergentism to substance dualism. I discuss Goff’s trajectory from emergentism to monism again in more detail below.

objects which future science could give us a transparent conception of. So microphenomenal properties (whatever it is like to be the most basic physical entities of which things are composed) will not constitute higher level facts about consciousness-as-we-know-it.<sup>60</sup> That we have very little grasp of what kind of phenomenal properties would be associated with fundamental entities implies that attributing experience to inanimate objects will be unenlightening as far as understanding our own experiences goes. If anything, constitutive panpsychism's easy accommodation of mind-brain identity, and the subsequent temptation to attribute qualia to inanimate objects, suggests it doesn't have a great deal to do with consciousness as we understand it. The kinds of intrinsic properties we attribute to particles, as whatever mediates between their structural and dynamic properties, may be "weird and wacky", but this wouldn't commit us to panpsychism.<sup>61</sup>

While always somewhat favourable to non-reductive forms of neutral monism<sup>62</sup>, Goff's criticisms of its reductive forms have at times led him closer to substance dualism. After all, it is difficult to read "layered monism" as anything but an oxymoron; o-properties supposedly 'float above' the empirical world of elementary particles,<sup>63</sup> and "are fundamental *entities*" (Goff 2015a, p390, emphasis added). Certainly ghosts were never intended as a rerun of earlier arguments, such as W.D. Hart's (1988/2007), who really does take the conceivability of experience persisting in the absence of a body as indicating a real possibility. Goff has never suggested that the mind could exist on its own (he thinks it would be hard to imagine a bodiless mind "out there in space" without ending up imagining a "floating sheet"<sup>64</sup>). He insists that they are not being defined as non-physical,<sup>65</sup> and has

---

<sup>60</sup> This receives its strongest formulation in one of Goff's earlier papers: "I conclude that a commitment to panpsychism does not help us to explain o-experience; o-experience being the very thing we want an explanation of consciousness to explain" (Goff 2009b, p310).

<sup>61</sup> *Ibid* p305

<sup>62</sup> Goff (2010) p129

<sup>63</sup> Goff (2015a), 397. "Against Funny Physicalism" goes into more detail on p.8.

<sup>64</sup> Goff (2010) p124

<sup>65</sup> *Ibid* 123

more recently argued that “The thing we end up conceiving of at the end of the second meditation is not just a thing with mentality not realized in physical stuff; it is a thing with mentality not realized in *any* stuff” (2014b, p7).<sup>66</sup> But this is unconvincing: if mental properties are such as to rule out their being identical with physical properties, and Russellian monism is assumed to be true, then the mental tells us what substance *couldn't* be (it couldn't be exhausted by properties which could not coherently be thought to bond into o-properties), and positively tells us what it *is*, namely, that substance in itself has an inner dimension of feeling. This is the first intimation of inconsistency in Goff's position; the wish to suspend judgement about matters of fact, give as much ground as possible to the sceptic, and, following Descartes, discover sense-certainty as beyond doubt, is in tension with the wish to do metaphysics, a sub-discipline which the convinced sceptic will deny is possible. And in fact Goff sometimes seems to proceed with doing metaphysics while denying that important progress can be made (see discussion of his 'noumenalism' below).

Consequently, the appearance of dualism arises primarily because Goff has never clearly answered the question of what o-phenomenal properties are supposed to be intrinsic properties *of*, if not the subvenient base of entities known to modern physics; after all, the latter is supposed to be ruled out once we have rejected constitutive Russellian monism. He maintains that the cause of phenomenal bonding must be non-empirical, in the same sense that qualia themselves are, and now postulates a “non-mathematico-causal” form of “spatial relations”, distinct from the familiar kind spatial relations into which empirical objects enter (2017b, p292-4). Alongside the distinction between intrinsic/discrete and

---

<sup>66</sup> In this paper he also denies that Cartesian doubt can give insight into the essential properties of *substances*, since their complete essence might consist of more than what is contained in the concept of either mind or matter. He has recently clarified that he considers himself a property dualist, and takes physicalism to be untenable because it cannot accommodate the kinds of *properties* that are revealed to be indubitable at the end of the doubting procedure (Goff 2014b, p12). This is a retreat from his earlier usage of scepticism in order to rule out non-mental aspects of one's essence (see n.56, above). I return to this issue in ch.4, but for present purposes the question of whether o-properties are fundamental substances or properties of fundamental physical substances is not an issue. The issue is how o-properties could be either of these in the first place.

extrinsic/relational, Goff therefore posits an additional distinction between extrinsic-relational properties and intrinsic-relational ones, which we can neither perceive nor imagine (he thinks doing so would be equivalent to having a solution to the problem of other minds<sup>67</sup>). And these intrinsic-relational properties could only be seen to “intelligibly” constitute o-properties from a God’s eye view. The impression that he is defending dualism only increases the more mysteries are invoked to explain how o-subjects come about in spite of their clash with the scientific worldview.

Russellian monism’s main advantage over dualism was originally that there is a direct analogy between facts about collections of physical properties as they are conceptualised scientifically, and facts about phenomenal properties; the two are supposed to share the same *structure*, and similarly (as Chalmers puts it) to manifest the same informational states,<sup>68</sup> even if their manifestations to consciousness are quite different. If the faculty of introspection and those of empirical perception could overlap, the correlation between qualia and brain states would look necessary, rather than contingent (see [1.2.2], above). But phenomenal bonding relations are simply not the kinds of properties to which there is any analogy in the physical world – or so Goff argues. If ghosts really reveal all there is to know about their own properties, then they cannot be identical to brains-as-we-know-them. Knowledge of the qualitative informational states with which physical states would be identified will therefore have to be, in some sense, *a priori*. So if Goff is right, introspection reveals an as-yet-unknown physical substance, and we are going to have to wait for science to catch up with his insights. This is a terribly immodest position for a metaphysician to end up with, and one not helped by the fact that he has not done a great deal to flesh out his alternative, non-constitutive form of panpsychism anyway. Instead, he has tended to focus

---

<sup>67</sup> Goff (2015a), p379.

<sup>68</sup> Chalmers has in mind the formal definition of information as a means of distinguishing a potentially meaningful signal from noise. See *TCM* p277-80.

on polemics against his opponents. And in fact there is a critical tension lurking in the way ghosts are invoked to either refute physicalism and constitutive panpsychism, on the one hand, or to form the foundation of Goff's constructive metaphysical project, on the other.

The polemical, negative role played by ghosts sees them defined as lacking any causal or functional properties whatsoever, since these have been doubted away, and as existing only for the 'duration' of the present moment.<sup>69</sup> Finding our selves to be beyond doubt does not seem to rule out Cartesian doubt about one's prospects of continued existence, or memories of the past; strictly speaking, 'I' might be completely transitory. Goff takes this hard line so as to refute the identification of conscious states with functional properties, the definition of which must make reference to causal interaction. Constituting macrophenomenal states according to constitutive panpsychism is likewise something *done by* microphenomena, which would presuppose the existence of causal properties. Making structure and dynamics intrinsic to the nature of consciousness, however, could be a slippery slope to it being nothing over and above its structure and dynamics.<sup>70</sup> The point of Cartesian doubt, which leads us to conceiving of a lonely ghost, is to sharply separate the phenomenal properties from their functions, e.g. pain's role in directing behaviour, in order to suggest that physicalism leaves experience out.

But conversely, if a ghost's properties are "exhausted by conscious experience" (Goff 2012, p745), then we do not seem to end up imagining something intuitively resembling ordinary or 'organic' subjectivity. Galen Strawson, in "Realistic Monism" (Strawson 2009) distinguishes between "the *thick* conception according to which it is only human beings and other animals *considered as a whole* that are properly said to be subjects of experience" and "the *thin* conception according to which a subject of experience...does not and cannot exist without experience also existing, experience which it is having itself" (p59). While this does

---

<sup>69</sup> Goff (2012), pp742–746.

<sup>70</sup> Heil disputes this, as discussed in chapter 2.

not amount to the same contrast as between macro- and micro-phenomenal consciousness – since for Goff even ‘thin’ subjects consist of many experiences bonded into one – it would be a weakness of Goff’s metaphysical picture if it cannot account for our ordinary sense of personal identity through time (we might recall William James’s reflections on experience’s temporal extension<sup>71</sup>), much as Chalmers’ inability to account for the feeling of subjective unity motivated scepticism about the adequacy of constitutive panpsychism.<sup>72</sup> This is not something that worries Strawson, who is happy to deny the existence of the self as anything but a bundle of discrete experiences (2009, p58), but then again, he is also happy to call himself a *constitutive* type of Russellian monist. Goff (2017a) admits that a plausible metaphysics of consciousness should avoid replacing our notion of personal identity with a string of unrelated, momentary subjects, though he prefers to avoid committing himself either way (269-70).

This raises the question: what else might o-properties be, if not lonely ghosts? If Cartesian doubt is a reliable way of drilling down to the indubitable, essential properties of things and reveals us to be no more than lonely ghosts, then so be it. It doesn’t seem to be the sort of thing Goff should be allowing himself leeway on. (This will turn out to be important in chapters 5 and 6, which will look in more detail at the question of how ambiguous a phenomenon consciousness is, and hence how much room it leaves for reinterpretation) After all, ghosts are supposed to supply the foundation to Goff’s positive alternative to physicalist and constitutive panpsychist metaphysics, which has been the focus of his more recent writings.

---

<sup>71</sup> See James (1983), pp 573-5.

<sup>72</sup> O-subjects have gotten even thinner since then, with Goff more recently giving ground to those sceptical of Cartesian certainty by allowing that only key experiences such as red are fully transparent at the end of the doubting procedure, rather than whole bundles (2017a, p3). I return to this issue in ch.5.

## [1.5] Goff in Focus II: Speculations

Goff's position on lonely ghosts can be contrasted to his positive metaphysics, which goes beyond the introspective evidence of sense-certainties in the present moment, and postulates phenomenal bonding properties which we are no more aware of from the first person than we are from the empirical scientific perspective. He compares his solution to the bonding problem to mysterianism<sup>73</sup>, a position often associated with Colin McGinn. McGinn argues that we are constitutively unable to make philosophical sense of natural properties which would account for mind-brain identity, in spite of the fact that these properties invoke no magic and may, in fact, be implied in some way by what we already know about the brain.<sup>74</sup> We are supposedly unable to infer these properties' existence, in the same way that birds are unable to infer the shape of the Earth despite its shape bearing directly on their migration patterns. McGinn notes that the distinction between the world we know, which he portrays as irremediably mysterious (hence 'mysterianism' as a label for McGinn's position), and the world in itself, is something like Kant's distinction between, respectively, phenomenal and noumenal,<sup>75</sup> wherein the latter is sealed off from metaphysical enquiry. And Goff uses the term "noumenalism" to identify any view that places full resolution of the mind-brain dichotomy off-limits to beings like us – views such as his own, in fact.<sup>76</sup>

Unlike McGinn's mysterianism, however, Goff's brand of noumenalism does not identify the entities that mysteriously produce consciousness with *empirical* brains, but with brains-in-themselves (*qua* intrinsic relations holding between their particles' intrinsic

---

<sup>73</sup> Goff (2017b), p294-5.

<sup>74</sup> See McGinn (1989).

<sup>75</sup> *Ibid*, p351, 358.

<sup>76</sup> Goff (2017a), p185.

properties). His book (2017a, p181) clarifies his position by arguing that in conceiving of objects in space we inevitably isolate them conceptually from one another; he concludes that conceiving of the relations between subjects in spatial terms is the cause of the bonding problem, as well as the problem of other minds. This is a helpful gloss on his (2017b) paper which first set out a solution to phenomenal bonding; Goff's mixture of diffidence and audacity is again on display there, in a crucial paragraph summarising his position:

[I]t is not surprising that we lack a transparent grasp of the phenomenal bonding relation – if such a thing there be – given the nature of our epistemic situation. Our most basic empirical science, physics, yields understanding only of the world's mathematico-causal structure, and the phenomenal bonding relation is not a mathematico-causal relation: conceiving of [micro] subjects standing in mathematico-causal relations does not remove their conceptual isolation, and hence does not remove their metaphysical isolation. Apart from its mathematico-causal structure, arguably the only feature of the world we transparently understand is consciousness. And consciousness is a monadic property. Our unfortunate epistemic situation does not afford us a transparent understanding of the (non-mathematico-causal) relations which conscious things bear to each other. (2017b, p292-3)

Here our inability to solve the mind-body problem is less a case of some contingent limitation in our cognitive apparatus (which psychology or neuroscience could conceivably uncover, or even fix), than a necessary upshot of our being finite beings in the first place. Not only would it take a godlike spectator to render Goff's postulated non-mathematico-causal properties perspicuous to itself; only a god could ever know there was another world of intrinsic properties to be discovered beneath empirical reality in the first place. Goff has put forward a whole universe lurking in the shadows of our own, complete with weird ("non-mathematico-causal") properties unknown to either science or philosophy. This is not simply



dualism between brains and ghosts, then, but between matter and what I suggest is best designated as the ‘astral plane’: the place where ghosts live.

### [1.5.1] Aristotle

The need to radically rethink relational properties in order to make room for phenomenal bonding is fairly constant across all of Goff’s writings. However, the way in which it is put into practice has varied. In particular, the relationship between the kind of consciousness we ordinarily think of ourselves as having, on the one hand, and Goff’s metaphysical speculations, is left unclear. Prior to *Consciousness and Fundamental Reality*, the closest he ever comes is in “Against Funny Physicalism” (2013, unpublished), in which he argues that to make “sense of entities that are both grounded [i.e. supervenient] and fundamental,” we should turn to Aristotle, for whom “organisms have parts, and yet are as fundamental – as metaphysically heavyweight – as their parts” (p11).<sup>77</sup>

It is easy to see why a return to Aristotle might appeal to Goff as an alternative to panpsychism. Burnyeat (2003) notes that for Aristotle, “the flesh, bones, organs, etc. of which we are composed are essentially alive, essentially capable of awareness” (p25-6), indeed, “already pregnant with consciousness, needing only to be awakened to red or warmth” (*ibid*, p19). These remarks are supposed to illustrate an essential difference between organic and nonliving material, rather than the presence of a lifelike awareness throughout the material world, as panpsychists claim. Whether o-properties could be integrated any better with the natural world on an Aristotelian model (his so-called hylomorphism) than a Lewisian one is a question I will set aside for chapters 2 and 3; for now, it is important to note the disparity between the way Goff usually conceives of

---

<sup>77</sup> Which might in turn presuppose a “thick” conception of subjectivity, to put it in Strawson’s terms; see above.

phenomenal bonding and what hylomorphism, the distinction between form and matter, was supposed to mean.<sup>78</sup> As discussed above, for Goff o-properties are qualia, conceived as impersonal objects inhabiting the supersensible realm of things themselves, which have been bonded by means unknown into a single aggregate. While the bonding relation is not supposed to be one we can visualise or conceive in mathematical or causal terms, the implication does seem to be that o-subjects are big objects of a special kind.

For Aristotle, the motivation for conceiving of organisms as essences (properly, substances, in his lexicon) comes from his desire to explain causation and the passage of time without falling prey to paradoxes which were exploited by philosophers at the time (e.g., Parmenides and Zeno) to deny there could be any such thing as change (Shields 2013, p57-60). Put briefly, the distinction between matter and form accommodates the thought that reality is unchanging, by postulating matter as that which underlies and persists through change, while introducing form as that which is gained or lost, thus explaining the manifest appearance of change. Moreover, Aristotle claims, “apart from things being changed, there is no change” (*Physics*, 200b32); the passage of time is numerable to the extent that natural kinds persist. And this requires the existence of substances, forms that are essential to making things the kinds of things they are, and are conceptually irreducible to more basic forms; the bearers and explainers of properties possessed ‘accidentally’. The most important point, however, is that, contrary to Goff, substances are not big things or even special relations between smaller things, but causes responsible for actualising matter as a determinate thing. In the case of organisms, the cause is simultaneously final, i.e. what they exist for, and substantial, i.e., what they are (Shields 2013, p106), since organisms are

---

<sup>78</sup> The next section discusses positions that Goff has occasionally defended which could be more accommodating of Aristotle, at least insofar as they build temporal extension and causation back into the metaphysics of mind; as I note in the current section, however, there are very serious clashes between Goff in his sceptical mode of argumentation, with appeals to lonely ghosts, and Aristotle. Goff’s positive metaphysical project both fails to recognise that clash and fails to develop a position that is compatible with either Aristotelian or modern scientific ontologies.

thought to exist for the sake of themselves. Functions supporting the life of the organism are of necessity temporally extended,<sup>79</sup> and it is these which contextualise, give form to, the aggregate of matter making up an organism at a given time.

Shields asks rhetorically whether we should “regard the group of 452,393,288 atoms as one thing and the form as another, to be welded together by some metaphysical blowtorch or other?” (2013, p311) – a picture heavily implied by Goff’s atemporal account of the bonding relation. No, Shields replies, because “the form and matter are not present as discrete, detachable entities. Rather, the form, as actuality, makes the proximate matter, as potentiality, an actual F thing”.<sup>80</sup> Matter, for Aristotle, only persists in actuality to the extent that it continuously gains and loses substantial forms. Considered in abstraction, a thing’s matter accounts for its being finite in magnitude and possessing the capacity to lose its current forms (both its current accidental forms, such as being chipped or being angry, and its substantial form); it plays a negative role as that which makes a substance something less than eternal; though the potential for change is in and of itself an unchanging property which all things possess. Hence, there is a sense in which there is no matter *in itself*; only enformed matter.<sup>81</sup> And this means that, if the life processes are essential to the definition of organic matter, then organic matter ceases to be upon death; a body only “homonymously” persists across the interval between life and death (Shields 2013, p98).<sup>82</sup>

For this reason, Burnyeat (2003) concludes that Aristotle’s worldview cannot be reconciled with our own: for Aristotle, “The unity of science is achieved from the top down, not from the bottom up, which is the way we have seen it since the seventeenth century.

---

<sup>79</sup> This point applies more generally to Aristotle’s analysis of causes, which are always contemporaneous with their effects; see Shields (2013), p77-8.

<sup>80</sup> See Aristotle, *Metaphysics*, 1041b11–31, where he argues for the incoherence of conceiving a substance as being realized by either one special element or as an aggregate of all its elements: in both cases the elements may persist in the absence of the particular substance they compose.

<sup>81</sup> This point will be discussed in more detail in chapter 3, including the question of whether Aristotle ultimately left room for un-mattered form.

<sup>82</sup> Don’t its elements persist? Yes, but this needs qualification; again, see chapter 3 below.

Aristotle simply does not have our task of starting from the existence of matter as physics and chemistry describe it and working up...” (p22). But that is precisely the task Goff faces. Moreover, he has not sought to modify his commitments to Lewisian metaphysics. Sider’s and Lewis’s ontologies deny fundamentality to organisms just as much as to furniture, and Goff agrees with them in thinking of science as reducing everything to point-particles.<sup>83</sup> So we are left wondering what the relation between o-subjects and their bodies could be. We are also left with problems for understanding mental causation, which turns out to be a more pressing issue for Goff than it was for Chalmers, and motivates his shift towards cosmopsychism; or so I shall argue.

### **[1.5.2] Mental Causation**

Goff’s views on mental causation flow from his aforementioned static or atemporal conception of composition, or bonding, which is decidedly un-Aristotelian. Goff concedes that “[t]he difficulties involved in making sense of the causal efficacy of non-fundamental properties are complex”, and that “overdetermination looms” whenever facts have truthmakers at non-fundamental levels of reality (Goff 2010, p131-2).<sup>84</sup> Mental states, for Chalmers, had causal efficacy under a physical description only on the assumption that those physical states could be constitutive of qualitative informational states. In light of Goff’s attack on constitutive panpsychism, then, he should be opposed to Chalmers’ epiphenomenalism, which affirms the apparent contingency of mental-physical correlation yet denies any hidden causal links. So his positive metaphysics leads to needing to credit ghosts with hitherto-unknown causal powers, in contrast to the way they are portrayed in Goff (2012), where they serve a negative and critical role, and are defined as ‘lonely’ ghosts

---

<sup>83</sup> Chapter 2 seeks to explore this missed opportunity in more detail.

<sup>84</sup> See also Goff (2015a), p396.

lacking causal powers or temporal extension. Otherwise, if ghosts are neither constituted by lower level goings-on, nor have detectable causal powers, what are they doing there? What physics are they mapping onto? Chalmers' property dualism rests on having some sort of relation between physical and mental facts - supervenience linking lower level causes to mental ones. But Goff denies that these sorts of relations can account for ghosts: only physical relations can be abstracted into bits. Hence the *kinds* of causes taking place at the non-mathematico-physical level can't be the kinds of causes known by mathematised physics.

Scientific evidence of causal effects irreducible to, or otherwise other than, relations between the grain of particles in motion would appear to be the only point of overlap between physical and mental states for Goff to avail himself of, if he has any sympathy for Russellian monism. He even says so explicitly in his (2017b), suggesting that "the proponent of phenomenal bonding might identify some empirically known relation as the phenomenal bonding relation" (p293). He argues that there should be evidence of powers "not...predictable on the basis of the behaviour of the system's parts. Hunting the phenomenal bonding relation, for the emergentist, will be a matter of looking for an empirically distinguished relation which relates the parts of systems with emergent causal powers" (*ibid* p296). But he goes on to reject this possibility in another paper,<sup>85</sup> on the basis that biological systems are not as sharply distinguished from non-living things as consciousness is from unconsciousness; and he denies there is any ambiguity in conscious experience.<sup>86</sup> He therefore seems to conclude that searching for emergent biological properties is a dead-end as far as understanding consciousness goes. This signifies the beginning of his shift away from dualism, towards cosmopsychism (though he still voices

---

<sup>85</sup> Goff (2014a).

<sup>86</sup> Although there is reason to think he is also inconsistent on this, as in his book he allows for the possibility of "confused" experiences which are to some extent opaque to us. See Chapter 5, where partially opaque experiences are discussed in more detail.

scepticism about reducing the causal powers of organisms to their parts in Goff (2017a), p252-3, regardless). But similar arguments against reductionism were made in Goff's earlier papers precisely in order to shore up the comparatively common-sensical position that organic life is metaphysically basic somehow (e.g. his appraisal of Aristotle in "Against Funny Physicalism", above). A naturalistic account of phenomenal bonding never seems to have been what he had in mind; what has changed over time is how much of common sense psychology Goff is prepared to jettison along with mainstream science.

In spite of voicing worries about mental causation, Goff does not situate himself clearly in the narrative of scientific reduction which is often taken for granted by metaphysicians, and which motivated constitutive Russellian monism in the first place. Goff allows that Russellian mind-brain identity is, in some respect, and so far as we can know, a brute and inexplicable fact. But he does not wish to defend the dualistic view of 19<sup>th</sup> Century so-called 'British emergentists', who thought the mind must spontaneously emerge from matter given the right conditions. Goff's most thorough critique of constitutive panpsychism (Goff 2015a) leans heavily on Brian McLaughlin's paper, "The Rise and Fall of British Emergentism"<sup>87</sup>, which narrates emergentism's fall from grace as chemistry, biology and physics at the turn of the century demonstrated that there was no room for the brute emergence of "kinds with the power to influence motion in ways unanticipated by laws governing less complex kinds" (McLaughlin 1992, p51). In keeping with his praise of Sider, then, Goff seems to side with reductionists who think biology is reducible to physics. But he also defers in the same paper to Terence Horgan's critical history of supervenience,<sup>88</sup> which argues, on similar grounds to McLaughlin, for the inability of this modern successor to emergentism to make adequate ontological room for mental properties, and Goff of course

---

<sup>87</sup> McLaughlin (1992).

<sup>88</sup> Horgan (1993).

denies it is capable of capturing grounding relations between micro- and macro-consciousness in Chalmers' panpsychism.

The result is that Goff is forced in to the uncomfortable position of having to deny that o-subjects or ghosts are either brutally emergent from micro-physical or micro-phenomenal aggregates (since this would clash with what established science tells us about how properties emerge), or supervenient upon them (thus ruling out scientifically-respectable, i.e. constitutive, emergence regardless). It is hard to see how panpsychism could maintain its elegance so long as reductionism is simultaneously taken for granted with regards to empirical objects, and yet indefensible with regards to mental properties. In response, Goff toys with the idea that the causal closure of the physical is false, arguing that fully deterministic causation in objects as complex as brains is an unfalsifiable supposition: "...physics aims to give a complete description of the causal workings of basic entities, in relatively isolated situations. There may be emergent causal powers of which physics remains silent".<sup>89</sup> As alluded to above, he also postulates an "intelligible" form of emergentism in which the truth-conditions for emergence could be specified, but admits that intelligible causation would *appear* inexplicable to us, which he recognises places him closer to Colin McGinn's mysterianism than Chalmers' panpsychism. But in any case, his examples of an intelligible form of causation are of a "crowd of wizards...whose angry activities bring into existence a demon" (Goff 2015a, p382), and of light coming into being following God's willing that there be light. It is questionable whether possible worlds in which these events could take place would be any more intelligible than those in which emergence is brute; moreover, Goff even describes them as "an ontological free lunch" (*ibid*), thus failing to take McLaughlin's intended conclusion fully on board.<sup>90</sup>

---

<sup>89</sup> Goff (2015a), p372; see also (2017a), p252-3.

<sup>90</sup> Goff (2017a) clarifies intelligible causation in terms of the following distinction, between "*constitutive grounding*, in which the produced is nothing over and above the producer, and *intelligible causation*, in which

Having cut the appeal to Aristotle included in “Against Funny Physicalism” when it was published as Goff (2015a), Goff began to place less emphasis on mental causation and emergence.<sup>91</sup> Thereafter his papers no longer take organisms ontologically seriously, and more recently he has conceded that science and philosophy could overturn “our folk notions of what it is for someone to be free, or for something to be solid, or for time to pass.”<sup>92</sup> In the process, he has gone from postulating a non-scientifically detectable vital spirit animating the body with non-deterministic will, to allowing that the bonding relation has nothing much to do with the body. Of course, if the bonding relation were to be identified by its causal powers, then we would have an unrestricted proliferation of scientifically inexplicable causation on our hands. It is perhaps for this reason that Goff eventually seems to have retracted his defence of free will. This is not to say that his metaphysics has since become less extravagant. On the contrary, if ghosts are not individuated via their (in principle) scientifically-measurable powers, there is nothing to prevent a Russellian from supposing that phenomenal bonding relations are ubiquitous, which Goff certainly does suppose. Goff (2017b) defends ‘unrestricted panpsychism’ as a consequence of being unable to render the metaphysical truth conditions of organisms in the language of particle physics, which leads to a Sorites-style paradox of when an organism begins or ceases to be.<sup>93</sup> This entails that there would be as many o-subjects coterminous with one’s body as there are

---

the produced is something over and above the producer” (150). The latter is also glossed as a form of layered monism (151). In addition to the example of God’s willing that there be light, Goff also lists the creation of the present by the past as a kind of intelligible causation. How the creation of o-properties might be likened to either of these examples is left unclear, but by this point Goff had rejected layered monism anyway.

<sup>91</sup> The themes re-emerge in (2017a), but are treated as challenges for a sophisticated monism to meet, rather than motives for adopting ‘layered’ monism (i.e., dualism).

<sup>92</sup> Goff (2014b), p3.

<sup>93</sup> See also Goff (2014a). He considers the following *reductio ad absurdum* for the possibility of biological conditions of possibility for the emergence of consciousness or o-properties: “For a conscious object composed of [e.g] seven billion particles, it is implausible to suppose that the psycho-physical laws are precise such that the removal of a single one of those seven billion particles could render it non-conscious” (2014a, p88). Goff’s argument is reminiscent of Peter Unger’s “I do not exist” (Unger 2009), who argues that, since it is undeniable that organisms are composed of a finite number of cells, it is easier to deny organisms exist at all than to try to solve the paradox that losing some of their cells would not destroy them (however, see next footnote). The following chapter argues against thinking of organisms so restrictively.



ways of grouping its parts, and of grouping those with other nearby entities (in fact Goff never quite explains why distance ought even to be factored in; he suggests there could be a bonded set of qualia made up of his teeth, the planet Venus, and one of his readers).<sup>94</sup> With this, however, Goff has departed from anything approximating our ordinary sense of what consciousness is. O-experience cannot really identify something holistic about one's organic existence if phenomenal bonding is unrestricted; indeed, it now seems to entail massive revisions to our self-conception. However, I will set this complaint aside for the time being.

Goff (2017a) succeeds in pulling his speculations on bonding into a coherent worldview which is, at least, satisfying in its simplicity. In chapters 6 and 7, he toys with brains or bodies being mereologically simple, but compositionally complex, entities, a position he calls 'fusionism' (2017a, p152), which treats the way things are put together as "distributional" properties or aspects of the individual in question. The fact that organic objects would still have to be somewhat vague on this view does not come in for renewed criticism this time, but Goff does voice criticisms along the lines I have been pressing him on above. Following a reappraisal of evidence against brute emergentism, Goff (tentatively; see below) concludes that it supports "Nomic Generality", the view that "There is nothing that happens in cells or bodies or brains that could not in principle be causally explained in terms of the laws that apply to particles and rock and planets" (2017a, p245). And "Nomic Generality supports *anti-emergentism*, where emergentism is understood as the view that fundamental properties emerge at specific levels of complexity between the micro and the cosmic levels, such as the chemical or the biological" (246).<sup>95</sup>

---

<sup>94</sup> Goff is likely to have taken his cue from David Lewis's defence of the unrestricted composition of objects in Lewis (1989), p211-3. Lewis complains that language and common sense must fail to carve reality at the joints since they admit of vagueness, whereas the number of composite objects in existence must have a determinate answer. Meanwhile, Lewis's conception of the universe as fundamentally composed of point-particles provides no guide as to what those particles compose. On balance, the notion that every part-whole relation is instantiated is less counterintuitive than mereological nihilism (such as Unger's approach to organisms), so Lewis opts for the former.

<sup>95</sup> Goff illustrates the same point more vividly in an earlier draft, where he writes:

In another change of heart, Goff also argues (chapters 7 and 8) that grounding o-properties in the non-mathematico-causal bonding of microphenomena would render experience opaque to us; there would be a crucial fact about consciousness, its holism or lack of grain, which appears fundamental but is actually the upshot of unknown physical powers. Goff's response to all these problems is to push beyond fusionism, into cosmopsychism, whereby only the universe as a whole is fundamental. The opacity of individual consciousness concerns its relation to this whole, rather than the way in which its parts relate, and so the opacity to us of non-spatial relations is harmless, with respect to what we know about our own properties and parts, since the unknown bonding properties are located outside of us. Microphenomena are always subsumed by o-properties, which are subsumed by the cosmos. Presumably there could still be any number of overlapping o-subjects, such as one comprising Goff's teeth and the planet Venus, as he suggested above; but at least any arbitrary way of counting subjects would be less fundamental than the whole in which they are subsumed.

Cosmopsychism also helps minimise worries about mental causation. Goff distinguishes between the 'Humean' metaphysics of causation, which treats physical laws as "brute patterns or regularities that obtain among concrete entities" (247), and the anti-Humean that "insists that such patterns must be *explained*, and does so in terms of the

---

What takes place when an o-conscious entity comes into existence, according to the fusionist, is a radical change in nature. A huge number of micro-level entities pass out of existence and are replaced by a fundamental macro-level entity; we go from a situation in which trillions of things at the micro-level are in the driving seat, to a situation in which a single macro-level entity is in the driving seat. It would be weird if that change in driver didn't show up empirically; if the brand new macro-level entity continues to make the micro-level run just as it did when it ran itself. We would be left with the sense that nature was conspiring to hide this radical change from us. (Chapter 7, "An Elegant Theory of Matter", 2014d, p16)

He concludes that emergentists should "deny either the *causal efficacy of o-consciousness* or *causal closure of the narrowly material*" (*ibid*), neither of which seem to be attractive options (bear in mind that even Chalmers, despite defending epiphenomenalism about phenomenal properties, granted causal efficacy to the mental vis-à-vis its physical properties; Goff is arguing that fusionist emergentists cannot even have that). In the published version Goff poses the same dilemma in terms of having to either denying fundamentality to o-properties (constitutive panpsychism) or grounding them in the widely material cosmos (cosmopsychism).

causal powers of certain fundamental entities” (*ibid*). Cosmopsychism denies that the causal powers of the mind are grounded in micro-level determinations, and attributes all fundamental causal powers to the universe as a whole. So Goff is still sceptical about current scientific powers of explanation. However, the distinction between Humean and anti-Humean metaphysics makes a stronger case than earlier appeals to noumenalism did: We are in a reasonably good position to weigh in on the causal powers of microphysical entities. If these are not the ground of physical laws, then science is stuck with ‘Humean’ explanations, i.e., it can make inferences based on statistical regularity, but cannot truly ‘explain’ what is happening. We cannot currently run experiments on the whole universe, or observe far-flung regions of spacetime with anything like the precision of fundamental physics. So, like the Humean, we would be stuck with ‘best guesses’ until the mythical future time when ‘all’ physical facts are known to us. Because the constant conjunction of lawlike relations at the micro-level is not fundamental, on this view, we could never rule out that it is merely coincidental.<sup>96</sup>

Of course this way of putting it still sounds like selective scepticism. Goff is indeed sceptical that Nomic Generality is true (see n.91, above), but his argument does not hinge on this; cosmopsychism is marshalled as the most plausible upshot of combining the truth of Nomic Generality with the irreducibility of conscious states to micro-physical entities and processes. What Goff seems to have in mind is something more like anomalous monism. This is the view that lawlike generalisations can be made about physical processes, and that mental states are token-identical with them, but that connections between states under a mental description are not lawful or predictable in the same way. Understanding why somebody deduced conclusions from premises, or turned on a light, will inevitably make use

---

<sup>96</sup> I return to the issue of Humean vs. anti-Humean metaphysics in the next chapter, where (among other philosophers) Gregg Rosenberg uses anti-Humean arguments to build a more sophisticated and science-friendly form of panpsychism than Goff’s.

of such notions as reason, belief, and desire, which are irreducible to concepts taken from the hard sciences. It would be unfortunate if Goff's speculations came down to defence of a position first formulated by Donald Davidson almost 50 years ago.<sup>97</sup> But cosmopsychism does put a new spin on the irreducibility of mental causation, albeit one which Goff does not develop in much detail, since it comes close to the end of the book. The thought seems to be that *what it is* one is choosing, and therefore what *explains* one's choice, will depend on what is going on in the universe at large. This way of looking at things is compatible with complete determinism at the micro-level, and avoids selective scepticism. Since the universe, according to cosmopsychism, is a self-causing mental entity, o-subjects enjoy autonomy from the micro-physical derivatively, by their subsumption in the universe.

The aforementioned may seem vague and speculative. Part of the problem is that Goff doesn't illustrate the notion of grounding in the cosmos with any examples, so I will supply one here instead. The following is lifted from Richard Tarnas (2006), a historian of ideas writing in defence of astrology; he conceives of

the universe as a fundamentally and irreducibly interconnected whole, informed by creative intelligence and pervaded by patterns of meaning and order that extend through every level, and that are expressed through a constant correspondence between astronomical events and human events.... In the perspective I am suggesting here, reflecting the dominant trend in contemporary astrological theory, the planets do not "cause" specific events any more than the hands on a clock "cause" a specific time. Rather, the planetary positions are *indicative* of the cosmic state of archetypal dynamics at that time.... Instead of the linear causal mechanisms of matter and force assumed in a Newtonian universe, the continuous meaningful coincidence between celestial patterns and human affairs seems rather to reflect a fundamental underlying unity and correspondence between the two realms—macrocosm

---

<sup>97</sup> Davidson, "Mental Events" (1970/80).

and microcosm, celestial and terrestrial—and thus the intelligent coherence of a living, fully animate cosmos. The postulation of a systematic correspondence of this kind implies a universe in which mind and matter, psyche and cosmos, are more pervasively related or radically united than has been assumed in the modern world view.

As for the relevance of causality in understanding astrological correlations, it seems that a fundamentally new kind of causality must be posited to account for the observed phenomena. Rather than anything resembling the linear mechanistic causality of the conventional modern understanding, what is suggested by the evidence is an archetypal causality that in crucial respects possesses Platonic and Aristotelian characteristics... (*Cosmos and Psyche*, 2006, p77-8)

Goff, for his part, is keen to distance himself from “new age” associations one might form about his philosophy (2017a, p254); “Cosmopsychism does not entail pantheism. We need not think of the universe as a supremely intelligent rational agent ... It is more plausible that the consciousness of the universe is simply a mess” (*ibid* p243). But this is evasive; after all, individual subjects only possess properties irreducible to micro-physics, according to Goff, to the extent that they derive those properties from the whole universe. Tarnas’s fixation on our relations to planets in the solar system is quixotic, and astrologers’ readings of archetypal significance into the workings of the universe are too vague to be candidates for ‘joint carving expressions’ by any standards (and not simply those of Ted Sider). But the broad idea of a “constant correspondence between astronomical and human events”, due to the former grounding the latter, and the attendant revisions to our concept of causality, does seem to be what Goff has in mind. Moreover, were cosmopsychism to be accepted, some vague form of animism, with more than a passing resemblance to Tarnas’s view, would become defensible: as the ultimate bearer of the powers of mental causation, even a messy universe would be apt to have the property of irrationality attributed to it. To the extent that

o-subjects have agency and may be praised or blamed, so too must Goff's cosmos.

Moreover, to the extent that it manifests any causal regularity at all, the temptation would be to read some minimal form of intent into the universe's activities, just as we do with each other.

It would be too easy to conclude that Goff's metaphysics is somehow unserious or guilty by association with New Ageism in spite of his protestations. This is because the problem of animism which cosmopsychism entails can be answered in large part by refusing to credit humans with animate properties either. Galen Strawson, for example, denies that human beings have free will in even a compatibilist sense<sup>98</sup>, along with personal identity over time<sup>99</sup>, and the existence of a meaningful narrative to our activities over the course of a lifetime.<sup>100</sup> If true, this would save us the trouble of having to attribute versions of those properties to the universe. And I have already quoted Goff (2014b) voicing scepticism about "folk notions" of freedom and temporal extension (see n.89, above). Animism is only one horn of the dilemma he faces. The other is that by entertaining scepticism about key parts of our folk psychology, Goff risks sliding into scepticism about the holistic nature of consciousness, which was his initial motive for dispute with constitutive panpsychism and, more importantly, with physicalism (which by Occam's razor would seem to be preferable to cosmopsychism, in any case).<sup>101</sup>

Goff concedes (2017a, p269-70) that revisionism about our temporal extension would be unwelcome and bear a large burden of proof, but the same complaint can be levelled at his own metaphysics, whose ontology is not even clearly defined. Key terms such

---

<sup>98</sup> See his papers, 'On the Inevitability of Freedom from the Compatibilist Point of View', 'On the Impossibility of Ultimate Moral Responsibility', 'Consciousness, Free Will, and the Unimportance of Determinism', and 'Free Agents', all in Strawson (2008).

<sup>99</sup> See his appendix to "Realistic Monism", in Skrbina (ed.), 2009, p57-65, on the "Sesmet theory of subjectivity".

<sup>100</sup> "Against Narrativity" (Strawson 2004/2008).

<sup>101</sup> I explore this complaint in ch.5 and the conclusion of chapter 6.

as “non-mathematico-causal bonding” and “intelligible causation” are invoked to do a great deal of work in patching everything together, yet their nature is said to be “noumenal” (i.e. unknowable) to us. Goff’s own metaphilosophical appeals to the virtues of simplicity and coherence with common sense should encourage us to look elsewhere before accepting his account of the nature of consciousness. By contrast, while his forays into Aristotelian vitalism were short-lived, and seemingly difficult to reconcile with our decidedly non-Aristotelian scientific worldview, they at least had a historical precedent, and deserved further development. Having given up on free will and thus (as I argued above) one plausible way of rendering “intelligible causation” actually intelligible to us, he has ended up with the worst of both worlds: an implausibly thin conception of personal identity and an implausibly extravagant and incomplete metaphysics.

Since it was the earlier metaphysics of o-subjects, free will, and neo-Aristotelianism which came closest to defending the reality of “the phenomenal properties we pre-theoretically associate with humans and other animals” (2017b, p295), at this point it is worth segueing into a review of empirically-minded philosophers critical of those who draw wild conclusions from so-called Aristotelian essentialism,<sup>102</sup> and who take it upon themselves to defend pre-theoretical common sense against such conclusions – ironically via a reappraisal of Aristotle. Goff and Chalmers’ presuppositions about physical reduction and the nature of knowledge of essential properties will then be put to the test. The latter both consistently pose the problem of consciousness in the most general terms possible, in order to defeat any possible objection (so much so that one commentator has described Goff’s

---

<sup>102</sup> See e.g. Hart (1988/2007), who reads Cartesian scepticism as a way of approaching essentialist metaphysics, whereby a thing is a basic substance if it can be conceived of independently of anything else (and we can conceive of a disembodied mind). Hart claims to find the same methodology in Aristotle (1988, p122). And see Cohen (1978), p387-8, who traces Hart’s version of Aristotelian essentialism to W.V.O. Quine; needless to say, none of this has much to do with Aristotle himself.

manoeuvres as “heavy handed”<sup>103</sup>). While the following exposition will similarly skate over themes deserving fuller treatment, the point is merely to demonstrate the resilience of the Hard Problem and, moreover, to begin to distinguish it from the way in which it is posed by arch-property dualists exemplified by Chalmers and Goff.

---

<sup>103</sup> Simon (2014), p106.







## Chapter 2:

### The Bonding Problems

#### [2.1] Problems and Prospects

As discussed, Goff's main divergence from David Chalmers is in his appeals to old-fashioned Cartesian scepticism, which, at least initially, were defended as being on firmer ground than Chalmers' appeals to zombies. He was also briefly led to a reappraisal of Aristotle, specifically the organic vitalism defended in Aristotle's *de Anima*, before heading in a much stranger direction. Goff's excursions into ancient metaphysics turned out to be short-lived, because this kind of organic unity is quite unlike the complex machine of modern biology; leading him to the apparent dualism of o-subjects, on the one hand, and bodies made up of particle aggregates on the other. This is really a dualism between pre-Enlightenment and post-Enlightenment concepts of the physical, as well as what counts as adequate knowledge of the physical, and Goff resolves the tension by relegating science to dealing with 'merely' relational, as opposed to real, intrinsic properties. This is in spite of endorsing Sider's metaphysical realism. Consequently, his (1) Cartesian epistemology, (2) Russellian metaphysics and (3) non-reductionism together form an inconsistent triad. In particular, while Goff's increasing denigration of common sense is faithful to his Cartesian point of

departure, the Cartesian drive towards greater precision is also what gave rise to atomistic metaphysics, which Goff seeks to refute.

These problems can all profitably be viewed as variations on the combination problem that Goff first diagnosed in Chalmers; in the end, Goff falls prey to the problem just as much as his opponents. Following in the wake of Goff's polemics on the subject, David Chalmers' recent article, titled "The Combination Problem for Panpsychism", has attempted to give a complete taxonomy of variations on the problem first diagnosed by William James, and concedes that it poses a serious challenge to metaphysicians; so much so, in fact, that he is unsure how they could ever overcome it (Chalmers 2017, p210-11).

The results of Chalmers' survey are reproduced in a table on the next page, showing the bonding problem divided first into three main problems: subject, quality, and structure combination problems. Answers to each of these problems are then found to have sub-problems of their own, but they all share common defects, many of which come down to making sense of introspection: what is the subject that introspects, and how much authority does it have with regards to experience? We appear to enjoy a revelation of certain properties which rule out constitutive panpsychism (*ibid* 190). Denying this revelation, and the conceivability arguments in support of it, gets us on a slippery slope back to physicalism, but the alternatives are just as bad: dualism or overdetermination, both of which haunted Goff's o-subjects as a result of the structural mismatch between them and micro-physics, and the decomposition problem, which his more recent cosmopsychism is confronted by. Neither Chalmers' preference for constitutive, combinatorial solutions, nor Goff's tendency towards noncombinatorial responses seems up to the task of meeting all these objections, let alone preserving our common sense intuitions about experience.

Main problems	Sub-problems	Problems common to all	Types of responses	Conclusions
<p>Subject combination problem: "how do microsubjects combine to yield macrosubjects?" (182)</p>	<p>Subject-<u>summing problem</u>: "given any group of subjects and any further subject, it seems possible in principle for the first group of subjects to exist without the further subject" (182-3)</p>	<p>Unity problem: How do microexperiences come together to yield a unified consciousness? (183)</p> <p>Boundary problem: "How do micro-experiences come together to yield a bounded consciousness?" (183)</p>	<p><b>Combinatorial responses</b></p> <p>.....</p> <p><u>Infusion</u>: "low-level subjects 'merge' or 'blend' or 'fuse' to yield higher-level subjects. After the merging, the low-level subjects no longer exist in their own right." (198)</p> <p>Problems: causal overdetermination, and "Extrinsically identical physical systems... might be conscious or non-conscious depending on the surrounding context" (fn.8 p200, fn. P214)</p>	<p>Conceivability argument against phenomenal bonding &amp; functionalist deflation of awareness; these combinatorial solutions look ad-hoc.</p>
<p>Quality combination problem: "How do microqualities combine to yield macroqualities?" (183)</p>	<p>Palette problem: "How can [a] limited palette of microqualities combine to yield the vast array of macroqualities?" (183)</p>	<p>Grain problem: "How do microexperiences come together to yield homogeneous macroexperiences, such as a homogeneous experience of red, instead of an enormous jagged array of distinct qualities?" (183)</p>	<p>Deflate the subject</p> <p>Problems: difficult to imagine phenomenal properties of which no-one is aware; this solution "may seem to require eliminativism about phenomenal properties or, at least, a reconception of them as properties of quite different entities." (197-8)</p> <p><b>Noncombinatorial responses</b></p> <p>.....</p> <p>Holistic identity panpsychism: macroscopic entities &amp; macro-experiences are fundamental (includes <u>cosmopsychism</u>)</p> <p>Problems: Decomposition problem: "How does macroexperience give rise to microexperience? ... how does a single subject give rise to multiple dependent subjects? How do macroqualities yield microqualities, and how does macroexperiential structure yield microexperiential structure?" (194)</p>	<p>Risk of slipping back into physicalism if appeals to conceivability &amp; revelation are ruled out.</p>
<p>Structure combination problem: "How does microexperiential structure (and microphysical structure) combine to yield macroexperiential structure?" (183)</p>	<p>Structural mismatch problem: "Macrophysical structure...seems entirely different from the macrophenomenal structure we experience" (183)</p>	<p>Awareness problem: "How do micro-experiences come together to yield awareness of qualities?" (183)</p> <p>Protoexperiential-<u>experiential gap</u>: "How can nonexperiences constitute experiences?" (184-5)</p>	<p>Emergent panpsychism: "base entities do not metaphysically necessitate the emergent entities...Emergent panpsychism has the great advantage of avoiding the combination problem. Strongly emergent entities and properties are best construed as fundamental entities and properties, not grounded in the base entities or in other entities" (192-3)</p> <p>Problems: resembles substance dualism; "unattractive choice between epiphenomenalism, interactionism, and underdetermination" (194)</p>	<p>"Scylla of micropsychism; Charrybdis of cosmopsychism" (Rosenberg 87-8)</p>

A seemingly plausible response is to argue that Chalmers and Goff are both wedded to an outdated picture of scientific ontology, one which is overly deferential to fundamental physics; recall Sider's scepticism about the existence of tables. Why not deny the propriety of building metaphysics on a base of point particles, rather than on functions, processes, or information? Why does the most real layer of reality have to be the smallest, most inert grain? Reductionism seems to be a poor fit even for Chalmers' original constitutive form of information-based panpsychism, as he concedes in his paper on the combination problem:

It is not easy to see how this line [reductionism] can work for a constitutive Russellian panpsychist. From the perspective of physics, high-level information structures are derivative aspects of a more encompassing and more basic narrowly macrophysical structure. We might expect that on a constitutive Russellian view, macrophenomenal properties would have this more basic structure rather than the somewhat arbitrary informational structure. (Chalmers 2017, p209)

In other words, if Chalmers really has a combination problem it is because he is committed to information being supervenient and derivative; in this respect it is "somewhat arbitrary". As discussed in the previous chapter, the same worry about the arbitrariness of groupings led Lewis to embrace unrestricted composition. Whereas if structure rather than matter is most fundamental, then the combination problem does not arise. Chalmers considers but rejects this line of argument, which he associates with fusionism and emergentism, because the latter both suffer from causal overdetermination, treating higher-level entities as being equally real to their microscopic truthmakers.<sup>1</sup> In putting things this way however, Chalmers

---

<sup>1</sup> Goff (2010) also worried about this (p130-1), though, as I argued in the previous chapter, he has yet to tackle the problem of mental causation head-on.

is simply restating his commitment to the Lewisian picture of supervenience built on sparse properties. I will suggest in this chapter that we should not underestimate physics' ability to confound our intuitions. It may be that there are versions of physical theory that reduce matter to structure but do not suffer from the problems Chalmers diagnoses.

Chalmers, for his part, does not seem to have made up his mind on where to stand about the prospects of solving the combination problem. He is critical of non-reductionist theories which seem to violate logical supervenience (see [4.3.2], below), but he is aware that the problem would be much easier to solve if the universe were not constructed out of distinct bits. Chalmers concludes his survey as follows:

On my view, the avenues that seem to be perhaps the most worth exploring are phenomenal bonding or quantum holism (to solve the subject combination problem), small qualitative palettes (to address the quality combination problem), principles of informational composition (to address the structure combination problem), and a somewhat deflationary account of awareness of qualities to tie all these aspects together. (Chalmers 2017, p210)

Ironically, an answer along these lines was supplied some years earlier by Michael Lockwood in his *Mind, Brain, and the Quantum* (1990). Chalmers is not unaware of Lockwood, but does not address his theory in detail. Doing so reveals the difficulty of meeting Chalmers' challenge to construct a solution to the bonding problem that does not lapse back into scepticism about phenomenal revelation.<sup>2</sup>

---

<sup>2</sup> Phenomenal revelation is discussed in more detail in chapter 4.

## [2.2] Lockwood's Quantum Consciousness

Michael Lockwood focuses on, and addresses, the palette problem and the grain problem by appeal to quantum mechanics (henceforth QM), which, he argues, yields a universe as rich and smooth as phenomenal qualities seem to be; he argues emphatically against conceiving of the universe as being fundamentally composed of bits. Lockwood's metaphysics is an answer to the structure combination problem, by arguing for isomorphy between the smoothness of experience and of the quantum universe, but he also attempts (on more or less independent grounds) to answer the unity and awareness problems by deflating the subject, and to close the protoexperiential-experiential gap.

Lockwood (1990, p225-236) argues that the 'many worlds' interpretation of QM is better thought of as a single, large universe whose various possible states (from our perspective) blur ambiguously into each other – an extension of the static 'block universe' view of time.<sup>3</sup> The appearance of determinacy, and thus of grain, is down to the "selectivity" of our perspective on it, a perspective which comprises a small subset, almost an infinitesimal slice, of the larger whole, leading to the commonsensical intuition that the ambiguity of physics is limited to extremely small scales. The response that there is at least grain at the level of atoms and larger objects, and thus a bonding problem at any scale relevant to brain functioning, can be met, Lockwood argues, by appeal to indirect realism. He embraces the criticism that, for a panpsychist, a brain surgeon would simply be observing a part of his or her *own brain* (1990, p159), arguing that inference to the best explanation is all our indirect access to things beyond our internal states affords. The brain surgeon's array of phenomenal experience might seem to *represent* a world of determinate objects in indirect relations with one another, but the array itself is nothing of the sort: experiences

---

<sup>3</sup> See Hoefer (2011).



blur ambiguously into one another, particularly along the time dimension (1990, p99-100). Lockwood argues that experience lacks grain because it involves “a degree of *averaging* over spatial or spatio-temporal regions” (1990, p237) existing in a state of superposition, straddling a region of the multiverse, whose sheer size and complexity also means it can accommodate a vast palette of experiences.<sup>4</sup> There is only a clash between experience and reality so long as what experience represents is taken at face value, whereas physics now suggests that the familiar determinate universe is an artefact of our limited perspectives on it; in reality, *everything* is in superposition, so bonding is the norm rather than the exception. Lockwood concedes that there is still a mystery of why only certain slices of the superposition directly affect the senses, while the rest has to be inferred from experiments. But he chalks the “arbitrary selectivity” (*ibid* p237) of our perspectives up to our remaining ignorance of QM; this amounts to a refusal to answer the decomposition problem which holistic positions such as his suffer from.<sup>5</sup>

There are limitations with Lockwood’s approach. His epistemology presupposes that the norms of everyday and scientific knowledge are compatible with indirect realism (*ibid* 298-9), which he concedes is still a matter of debate.<sup>6</sup> And while the ontology of fields rather than particles goes some way towards solving the quality and structure combination problems, the selectivity of our perspectives on the universe leave a crucial strand of the structural mismatch problem unaccounted for, namely the clash between *subjects* of experience (as opposed to experience itself) and the apparent homogeneity of the universe

---

<sup>4</sup> Lockwood’s main dispute with Paul Churchland’s reduction of colour experience to brain functions is that their properties, as measured by neuroscientists, do not comprise a large enough state space to accommodate the range of experiences we enjoy, and so cannot be candidates for the Russellian identification of intrinsic, phenomenal properties with extrinsic structures and processes in the empirical world. (Lockwood 1990, p172-6) He concludes that neuroscience is at best measuring a slice or subset of the ‘quantum brain’. See fn.9 below.

<sup>5</sup> I discuss positions which pointedly refuse to answer some or other aspects of the mind-body problem in more detail in ch.4.

<sup>6</sup> I defend Lockwood on this in ch.6.

as such.<sup>7</sup> As mentioned, Lockwood's response to the latter is deflationist. At times, he seems to take a mysterian line with regards to subjective perspectives. He admits to having no idea "how or why certain brain events should conspire to form integrated perspectives, illumined by awareness" (293), and, in spite of arguing for a connection between the latter and the existence of meaning and normativity, also admits that "the question what *in general* it is to have a concept... seems to me to defy philosophical and scientific analysis alike" (311). This is close to admitting he has no satisfactory theory of consciousness to offer.

However, the thrust of Lockwood's case tends to make the more substantive claim that the nature of subjecthood is illusory, as opposed to unknowable, with the attendant risk of slipping back into physicalism. He likens the selectivity of conscious awareness to a searchlight illuminating parts of the brain (again, construed as an object extended both in time and in the space encompassing all possible particle trajectories, the latter having been incorrectly popularised as the multiverse):

To the extent that we do have a transparent grasp of the concepts that we bring to bear on our mental lives, those concepts may be seen as capturing certain intrinsic attributes of brain states. To the extent, however, that they are topic-neutral, they represent no obstacle to an identity theory anyway. Moreover, *this goes for the concept of awareness itself*. For it seems to me that we cannot be said to have a transparent conception of awareness. (Can one see the eye with which one sees?) To return to the searchlight analogy, what we see are the objects that the searchlight illuminates for us. We do not see the searchlight. Nor do we see the light: merely what the light reveals... If that is right, then it follows that there can be

---

<sup>7</sup> Lockwood actually seems to build this problem into his theory when he writes: "the contents of consciousness correspond to eigenvalues of a set of *observables* which, again, are distinct from anything that the physiologist is likely to settle on: the dissonance... might be compared to a block of wood, distinct cross-sections of which can reveal strikingly different sorts of patterns, depending on the angle at which it is sliced..." ("The Grain Problem", 2015, p159). Again, this amounts to a refusal to answer one of the variations of the bonding problem.

nothing in our concept of awareness, such as it is, that could debar us from identifying awareness with some kind of physical process in the brain — albeit that it remains profoundly mysterious, in physical terms, what form such a process could possibly take. (169)

In other words, Lockwood cedes the problem of individuating o-properties, which so vexed Goff, to the physicalist, via their identification with the brain's capacity to form phenomenal judgements.<sup>8</sup> Worse, insofar as the unity of consciousness rests on the foreclosure of awareness as to its true nature (namely, as a kind of brain process), some or all of the properties disclosed to awareness could also turn out to be illusory. For example, in light of the difficulty of squaring the contents of consciousness, conceived as a superposition of “compatible observables”, with the parts of the brain as it discloses itself to the physiologist (“The Grain Problem”, 2015, p157), it might seem easier at this point to declare the unity of consciousness itself illusory. From the side of things in themselves, then, there would be no true selectivity of the intrinsic physical properties going on, or any present moment selected from 4-dimensional time. This need not rule out our enjoying a revelation of phenomenal properties, of course; anti-physicalists would simply have to appeal to particular, paradigmatically vivid experiences rather than making claims about the total contents of awareness at any particular time, since these could turn out to be fundamentally ambiguous.<sup>9</sup>

Subject-deflationism runs throughout Lockwood's book. In spite of later linking the exercise of rationality to the meaning-bestowing powers of awareness, Lockwood states early on in *Mind, Brain, and the Quantum* that his main dispute with the eliminativist Paul Churchland is over qualia, not folk psychology: “beliefs and desires might turn out, strictly

---

<sup>8</sup> Section [2.2.2] goes into more detail about why Goff was right to think that this is problematic, even if his argument against it in Goff (2014a) overemphasised the Sorites paradox of composition.

<sup>9</sup> In fact, this is precisely the strategy upon which Goff has recently fallen back, a concession I discuss in more detail in ch.5.

speaking, not to exist” (1990, p19-20). Similarly, as we saw in the previous chapter, Goff was ultimately prepared to let science redefine or eliminate free will. And Chalmers took a functionalist line with regards to the nature of intentionality, making meaning and reference essentially a matter of the right kind of lawlike relations with the external world. His paradox of phenomenal judgement, meanwhile, which he answered by embracing epiphenomenalism, downplays the relevance of introspective awareness as well. So despite eliminative materialism featuring as something of a bogeyman for all three of these panpsychists, it must be borne in mind that their own accounts can be highly revisionist as well.

### **[2.2.1] Lockwood’s Mind-independent Minds**

The centrepiece of Lockwood’s subject-deflationism, and consequent proximity with physicalism, is his “disclosure view” (1990 p162) of phenomenal qualities. Self-consciously echoing the indirect realism of Descartes and Locke, in which the mind is only directly acquainted with its own properties, Lockwood identifies as a “naive realist” with respect to basic sensory experience, which he thinks of as properties internal to the brain<sup>10</sup>:

The naive realist with respect to material objects holds that we have an immediate and transparent acquaintance, in perception, with certain material objects and certain of their attributes. Furthermore, he holds that the objects and attributes which are presented in perception are in no way dependent (or their existence on being perceived; nor is there any mystery about what objects are like when we are not perceiving them. Even when

---

<sup>10</sup> This qualification is important in light of arguments for the theory-ladenness of observation; so-called holistic theories of meaning, often associated with Wilfrid Sellars and Donald Davidson. Against these philosophers, Lockwood urges that whatever conceptual scheme we bring to bear on experience will presuppose a foundation of transparently-grasped properties in order to “lend substance to the scheme as a whole” (1990, p167). An important twist on Lockwood’s argument here will be considered in chapter 6.

unperceived, they are essentially as they appear when they *are* being perceived (under favourable conditions). Perception serves merely to *disclose*, in part, the inherent nature of things... According to this view, phenomenal qualities are not qualities *of* awareness. On the contrary, awareness is *of* them". (Lockwood 1990, p162)

Lockwood again turns to the searchlight metaphor of awareness, arguing that what introspection *discloses* are objectively existing qualities of the brain which in themselves outrun awareness (*ibid*, p163). He gives the example of seeing a polka-dot shirt, arguing that there is an objective fact about the number of dots perceived even in the event that the shirt is a hallucination. This is not as far-fetched as it sounds. Ned Block has defended a variation of the claim that phenomenology exceeds conscious access – his “overflow argument” – in several papers, drawing on empirical evidence from the cognitive science of perception that has emerged since the writing of Lockwood’s book.<sup>11</sup>

Goff classifies Lockwood as a panprotopsychist without further elaboration.<sup>12</sup> But the disclosure theory is not suggesting that something neutral or presubjective *becomes* phenomenal under the introspective searchlight. Lockwood asks us to imagine that qualities such as green could exist exactly *as they are*, even when no-one is looking. Lockwood’s position is better understood as panqualityism (Chalmers 2015, p271-5). Chalmers points out

---

<sup>11</sup> See especially Block, "Consciousness, accessibility, and the mesh between psychology and neuroscience" along with subsequent commentary from critics and Block's replies, Block (2007), and Block (2011). Phenomenological overflow demonstrates the opposite effect to examples of conscious gappiness, such as those cited by Dennett (1991); Block appeals to it to show that consciousness is richer, rather than poorer, than introspection assumes. He argues that cognitive ‘blurriness’ explains change blindness (e.g. failing to notice a passing gorilla while watching a basketball game) as well as accounting for what was thought to be the gappiness of the visual field, rather than indicating, as it did for Dennett, that we suffer from a ‘hyper-illusion’ of experience seeming to seem rich when what really appears is sparse. Memory might not gestalt the differences explicitly, but they are still experienced pre-reflectively, proving that there is more to experience than reports about it. Phenomenal overflow fits with metaphysical essentialism – what consciousness is, is independent of our methods of cognitively accessing it. Block is keen to distance himself from verificationism about consciousness, arguing (2007, 486-7) that inference to the best explanation is as far as the empirical data will allow us to go.

<sup>12</sup> Goff (2009b), p292. Goff may be right that this position at least shares the obscurantism which he sees in neutral monist and protopsychist positions, even if it is not strictly identical to them. See below.

that the term was coined by the physicalist Herbert Feigl (1960), but, importantly, identifies Bertrand Russell (1921) as an earlier exponent of it as well. Lockwood concurs, noting that it was a live issue among phenomenologists during the first half of the 20<sup>th</sup> Century as to whether sense data could exist unperceived.<sup>13</sup> Moreover, as Chalmers notes,

Qualities are not identical with phenomenal properties: when redness is presented to me in experience, I have a phenomenal property, but I need not be red. Instead, we would intuitively say that I am aware of redness, and that phenomenal properties involve awareness of qualitative properties. Likewise, phenomenal properties are always instantiated by conscious subjects, but qualities need not be. We can certainly make sense of the idea of a red object that is not a subject of experience. (Chalmers 2015, 271-2)

For this reason, Lockwood avoids identifying qualities with qualia, which “are ‘directly intuited’, ‘purely subjective’, and known completely and beyond possibility of error, because for them seeming is being” (Lockwood 1990, 171). He hopes thereby to “halt [the] slide into panpsychism” (*ibid* 170), which he associates, uncharitably and inaccurately, with the ubiquity of cognitive or intentional relations to experience. Chalmers and Goff both reject panqualityism (see Goff 2017a 160-1), on the basis that by separating awareness from qualities it simply reintroduces a variation of the explanatory gap. We can imagine zombie worlds with qualities but no subjects. Goff also wonders how qualities could exist unperceived, anyway.

Lockwood may be ahead of his time; Peter Langland-Hassan has argued that a recent case of Siamese twins conjoined at the brain, who each claim to be aware of some of what

---

<sup>13</sup> “G. E. Moore...who first coined the term ‘sense-datum’ around 1910, remained to the end of his life open-minded on the question whether there could be sense-data of which no one was conscious, as also on the question whether they might have attributes beyond those they appeared to have. C. D. Broad judged it likely that there were unsensed sense-data, and that sensed sense-data were probably differentiated beyond our powers of conscious discrimination (Broad, 1919, p. 218; 1969, p. 265).” (Lockwood 1990, p170).

the other is thinking and feeling, supports revising the idea that all directly perceived mental items are necessarily one's own.<sup>14</sup> But Chalmers urges caution on this issue, arguing that distinguishing phenomenal properties from subjects altogether is a slippery slope to eliminating them (Chalmers, 2017, p197-8; see also 202-3); his proposal was for a "somewhat deflationary" account of awareness, not full-blown deflation. He maintains that phenomenal properties ordinarily presuppose an awareness of them,<sup>15</sup> and that they must at least have bearers, "which might then be taken to be subjects" (*ibid*).

There is another, closely related, position with which Lockwood may be compared: panexperientialism. Chalmers regards this term as synonymous with panpsychism,<sup>16</sup> but others are careful to distinguish it. The concept can be traced back to Alfred North Whitehead, who distinguished the term "consciousness" from "experience" in his *Process and Reality*.<sup>17</sup> He characterized the latter as a kind of pre-reflective emotion or feeling (1929/1957/1978, p162-3), and voiced doubts about the overall relevance of consciousness by comparison (*ibid* 308); "consciousness presupposes experience, and not experience consciousness" (*ibid*, 53). He identified consciousness with a relatively sophisticated exercise of thought or reflection. And Lockwood follows Whitehead's terminological distinction between consciousness and experience.<sup>18</sup> Inspired by Whitehead, and echoing Lockwood's misgivings about being labelled a panpsychist, Gregg Rosenberg (see [2.5]) identifies as a panexperientialist on the grounds that some panpsychisms build cognition or self-

---

<sup>14</sup> Langland-Hassan (2015). Lockwood frames his theory as a response to the similar neurological puzzle presented by split brains (1990, ch.6).

<sup>15</sup> See Chalmers (2013).

<sup>16</sup> Chalmers (2015), p247.

<sup>17</sup> Whitehead (1929/1957/1978).

<sup>18</sup> "Let us stop talking...about streams of consciousness, centres of consciousness or even selves or minds or persons. Instead, let us talk in terms of *experiences*. 'Experience' here is to be understood in the philosopher's slightly technical sense of a conscious state, happening, or sequence of states or happenings, that is experienced *as a whole*...." (Lockwood, 1990, p87)

consciousness into fundamental reality<sup>19</sup>; by fundamental mind he means something much more primitive.<sup>20</sup> Steven Shaviro, who again credits Whitehead, goes further and asserts the existence of “nonconscious experience”<sup>21</sup>, deliberately stretching the mental/non-mental distinction to breaking point in the hopes of thereby naturalising the mind. Rosenberg, meanwhile, is noncommittal about how far the concept of experience may be extended beyond what we can imagine, particularly with regards to its possible independence from cognition of any kind (Rosenberg 2004, p94). Keeping in mind Chalmers’ misgivings over separating experience from the conscious subject, Lockwood’s position may therefore be thought of as panprotopsychist, which was how Goff interpreted it in the first place.

Since, by awareness, or attention, Lockwood evidently means something like reflective or cognitive consciousness, we might profitably identify phenomenal qualities with Whitehead’s ‘experience’, in the broad sense that they are pre-reflective and, one might say paradoxically, mind-independent mental properties. It may be objected that this goes against the spirit of the disclosure view, since pre-reflective experience is a much murkier thing, liable to be transformed by being brought into conscious awareness, and thus unresponsive of the direct realist view of sensations Lockwood defended in (1990), p162. But these remarks need to be seen in light of criticisms elsewhere in the book against direct realist epistemology (296-7), where he insists that the Cartesian and the direct realist are “in the same boat” insofar as *what it is* that is directly present is “not self-intimating”. In light of his accommodation of sense data “differentiated beyond our powers of conscious discrimination” (n.13, above), his attacks on qualia construed as immune to error, and his

---

<sup>19</sup> For examples of pan-cognitive variants of panpsychism in ancient and modern philosophy, see David Skrbina’s survey, “Panpsychism in History: an overview”, in Skrbina (ed.), *Mind that Abides* (2009), pp1-32.

<sup>20</sup> “we should distinguish between experiencing and consciousness, where the former would represent a kind of raw experiential protoconsciousness existing broadly in nature and where the latter names a species of experiencing shaped specifically by cognition to have certain special attributes we would associate with consciousness (such as representational meaning and awareness).” (Rosenberg 2017, p171)

<sup>21</sup> Shaviro (2014), p79-80



insistence that the ontology of experience is both “a substantial matter of fact” (171) and open to (mis-) interpretation (277), we must conclude that Lockwood’s qualities are susceptible to error through misidentification after all. If the independence of phenomenal qualities from awareness is, as Goff argues, just too implausible to take seriously, then panexperientialism might serve as a more charitable reading of Lockwood’s position. The latter has no commitment to unsensed qualities existing exactly as they are perceived, so long as perception, awareness, or consciousness, is understood as a partially cognitive act of interpretation.

I will suggest that the slippage from Chalmers’ identification of non-cognitive experience with what we ordinarily mean by consciousness or *psyche*, through Rosenberg’s more nuanced usage, and eventually to Shaviro’s paradoxical use of the term to denote something almost entirely foreign to introspection, is almost inevitable once a deflationary account of the subject is followed through. And since deflating the subject is, as Chalmers suggests, a promising step to saving panpsychism from the subject combination problem, the subject-summing problem, and the unity and boundary problems, its failure to account for the kind of consciousness we think we have (as opposed to a kind of ‘experience’ which we could never know we have) reflects unfavourably on Russellian monism as a solution to the problem of fitting qualia into the natural world.

These problems with the disclosure view all point to an issue with its purported nonreductionism, which is that it offers a much cheaper response to the bonding problem than quantum holism, so Lockwood’s solution to the mind-body problem is overdetermined. Once we have allowed for a disjunction between subject and experience, there is no need to posit quantum holism as a solution to the grain problem, when a blurry Cartesian theatre could do just as well in foreclosing the appearance of grain. Depending on one’s choice of metaphor, the inner eye could be long-sighted, the spotlight could be too dim, etc. (This is

not to say that quantum holism is a purely a priori posit; just that, to the extent that empirical matters weigh in its favour, it does not get us any closer to a solution to the combination problem than the Newtonian 'billiard ball' ontology already permitted, so long as it still entails a separation between subjectivity and experience.)

Even if Chalmers is right that it is a conceptual fact that qualia have subjects, with the consequence that no sense can be made of experiences of which nobody is aware (Chalmers, 2017, p197-8), a distinction between pre-reflective selfhood and phenomenal judgements could explain away the subject bonding problem. Chalmers might be read as endorsing that possibility in *TCM* when he draws the following distinctions:

“The cognitive contents of perceptual states... will be carried by first-order registrations rather than by first-order judgments. A first-order registration need not be a state that is endorsed by the subject, but it is nevertheless a contentful state that is available to the subject and that plays a role in the cognitive system... We can think of first-order registrations as the immediate product of perceptual and introspective processes, before they are rationally integrated into a coherent whole” (*TCM* p232-3)

Chalmers notes that first-order registrations can be contradicted by their corresponding judgements. There are cases such as the Müller-Lyer illusion of two lines falsely appearing to be of unequal length, where judgement is a more accurate guide to what we see than our first-order registrations; conversely, one can miscount the number of objects one sees. Might bonded o-properties then be just a matter of “rational integration” rather than a problem for ontology? The real subject need not be as richly phenomenal as Goff's ghosts.

Goff's arguments against constitutive panpsychism initially seemed to resist this line of reasoning, since he insisted that the subject making the phenomenal judgement 'I think

therefore I am' is itself a phenomenal experience in which individual phenomenal properties (e.g. each dot on a polka-dotted shirt) are grounded. But as his concept of bonding has drifted away from our common-sense notion of selfhood, and begun to encompass exotica such as unrestricted bonding (one o-subject for every possible combination of particles in the universe), the relation between o-subjects and phenomenal judgements has likewise drifted apart. (In any case, Goff's arguments for why second-order phenomenal judgements must at least *tend* to be reliable are quite weak, and subject to empirical refutation – see Chapter 5 for more detail). This is not to say that the unity of subjecthood itself is just an illusion generated by second-order judgements. Perhaps subjectivity, construed as the possession of a point of view on qualitative properties, should be treated separately from the issue of o-properties, which in Goff's ontology are unified sets of phenomenal qualia.<sup>22</sup>

Another line of attack would be this: the bonding problems supply independent reasons for thinking that spatial, causal or mathematical relations alone could never group a set of physical objects, let alone phenomenal qualities, into one (see Goff 2017b, p287). Insofar as the problem of situating o-properties within the physical world is of a piece with delimiting the types of composite objects there are, it seems as if the unity of subjecthood, whether or not this unity grants us infallible knowledge of what our qualia are, supplies good reason for thinking that there is more to physical composition than either mereological nihilism or unrestricted composition. According to this line of thinking, it is Goff's attachment to Lewisian ontology which needs to be challenged, not the reliability of second-order judgements.

Unfortunately, Lockwood's alternative to Chalmers and Goff still falls short, despite going beyond Lewis's 'grainy' universe. Lockwood has clarified, since writing *Mind, Brain, and the Quantum*, that "no literal sense can be attached to the notion of the conscious mind

---

<sup>22</sup> I return to this retreat from Chalmers' and Goff's qualia-first approach in ch.6.

being distanced, in this fashion, *from itself*" ("The Grain Problem", p149). The 'blur' exists only in the mind's eye, and so cannot be called upon to explain the lack of grain in experience itself, on pain of infinite regress. But it is obscure of him to then still insist that awareness itself is "an emergent phenomenon" (151), in the sense of supervening on fundamental properties, without thereby losing its direct relation to what we are aware of, assuming Russellian monism is true. None of this would be an issue if there were a straightforward coincidence between the subject described by his metaphysics and the properties of first-order experience itself, as uncovered at the end of Cartesian doubt. But his quantum ontology is singularly unsuited to this task, as Gregg Rosenberg argues:

Interaction, although structured, is seamless in Schrödinger's world, and ... [we need] a reason for stopping it here, where there are human cognitive systems in one eigenstate. There does not seem to be a compelling reason to think Lockwood's proposal would result in anything less than a many-worlds-sized individual.... for Lockwood's view to work, we need to find a basis for the existence of an intrinsically preferred set of quantum mechanical observables at precisely the level at which awareness emerges. This, however, is just the boundary problem [of how subjects are individuated] rearing its head (Rosenberg 2004, p89)

As the subtitle of *Mind, Brain and the Quantum: The Compound 'I'* indicates, Lockwood does not seek to solve the boundary problem. Lockwood's ontology replaces the self-contained Cartesian subject with an indeterminate number of more-or-less overlapping subjects, similar to the unrestricted bonding Goff defended prior to his shift to cosmopsychism. It is then no longer clear what the experiences of this many-worlds-sized 'compound' individual have to do with one's first-person awareness. This is also surely a problem for Chalmers'

original information-based panpsychism as it was defined in *TCM*, since he must allow that informational states are vague (see the quote from Chalmers 2017, p209, above).

### **[2.2.2] The Sceptical Wedge**

Ultimately, the same distance between specific ontological claims made by scientists about physical reality and their relevance to solving the bonding problems is present in these panpsychists' claims about metaphysical reality, and for the same reason. Their positions fall out of logical analysis of the kinds of properties reality must necessarily have (e.g. intrinsic properties in addition to relational ones), independently of whether subjects are around to experience it. By submitting mental properties to the same argumentative strategy, these philosophers have already presupposed that the mind can be thought of objectively, as 'mind-independent'. That reality, for panpsychists, contains an irreducible element of subjective feeling, is arrived at by way of inference to the best explanation, rather than being known with Cartesian certainty: we are asked to believe that it would simply be too much of a coincidence if the nonrelational, ineffable qualities that realize physical entities had nothing in common with the qualities realizing our experience.

Importantly distinct from all this is the strategy of Cartesian scepticism, which makes no claims (at least prior to Descartes' appeal to God) about reason's power to know things as they are in themselves, and makes essential reference to the subject who doubts; everyone must 'discover' the *cogito* for themselves. But this discovery merely amounts to posing the mind-body problem; in seeking to solve it philosophers must eventually fall back on the sorts of methods appropriate to claims about mind-independent reality, via the analysis of concepts, empirical considerations, and inference to the best explanation. Consciousness seems to be of a different order to objective reality as we know it, but fitting subjective and

objective back together is nevertheless going to involve thinking about their unity objectively somewhere down the line. The very possibility of doing metaphysics, then, sets philosophers on a slippery slope to conceiving consciousness in terms far removed from what common sense might suppose Cartesian doubt to have revealed.

A variation of this sceptical wedge is exploited by Papineau (1993) to argue that the very possibility of thinking about conscious states objectively undermines the existence of an objective difference between conscious and non-conscious states. He targets the attempt to identify conscious awareness with specific processes or properties of the brain, and in particular, with any self-scanning mechanism. Papineau notes that any candidate for self-scanning which we discover will be an ambiguous object or process (whose extent, or even existence, is a matter of interpretation), just as 'experience' was for Lockwood, thereby requiring another, second-order circumscription by awareness. One cannot both construe awareness as a property which disambiguates one subject from another, and as an objective property out in the world, as the latter property will never be disambiguated *enough* to do the job metaphysicians ask of it. Insofar as the presence of consciousness is neither ambiguous nor dependent upon one's interpretation of the evidence, then, these physical features could not stand in for consciousness, and so cannot aid Russellians who want to match consciousness with some physical structure. If Papineau is right about this, then Lockwood's theory of consciousness is truly a dead end: not only are the eigenstates of QM too large, and too blurry, to compose o-properties, but the process by which conscious awareness picks out or 'discloses' those states under a "searchlight" cannot supply the much-needed determinacy either. Goff's insistence that consciousness be unambiguous, and his complaints about organic vagueness (Goff 2014a), are therefore also vindicated; at the very least, anti-physicalists must provide criteria for individuating subjects in nature if they wish to offer an alternative to physicalist metaphysics.

Papineau argues that, besides the obvious fact that it rules out animal consciousness (1993, 125), it isn't clear that features common to all forms of self-monitoring could be found (126-7), meaning we have no firm criteria for what self-monitoring really amounts to. It's not clear, for instance, that we introspect emotions *in the same way* as sensations; and even if we did, being able to pick out the properties of introspection in our own cases would be of no help in deciding, from the third person, whether aliens, machines or intelligent animals introspect in the right way. He considers a dispositional account whereby states count as conscious "if they occurred in beings who could introspect, imagine and so on" (127), but wonders "what is supposed to stop us considering super-trees, say, or super-stones?" (128). Nothing, say the panpsychists – but they must still distinguish somehow between conscious awareness and experiences of which no subject is aware, if their position is not to recreate the explanatory gap in another form.

Moreover, on the subject of sensations Papineau considers the panexperientialist distinction between experience and consciousness, remarking that it "wouldn't matter too much if there is no principled basis for deciding whether fish are conscious, if there is a fact of the matter on whether they feel pain" (129); in other words, one might still be able to identify "such specific conscious states as pain or colour experience" (ibid) even in the absence of a general marker for whether they were 'disclosed' to any particular subject. But Papineau goes on to argue that this fails to account for the link between sensations and functional roles, which it is normal to assume in order to avoid chauvinism as to whether alien creatures could feel pain. Where the connection between the two comes apart, e.g. in imagined cases of inverted colour spectra, or "mad pain" whose awfulness, for madmen, compels them to snap their fingers and perform arithmetic, then we are once more at a loss to say *what* experience was taking place, with or without a subject's awareness. Papineau

concludes by “reject[ing] the intuition that there is a sharp line between conscious and non-conscious states” (124), which he regards as the basis for objections to physicalism.

In summary, Lockwood’s approach fails because what it does tell us about the physical structures with which consciousness is to be identified clashes with the feeling we have that experiences are circumscribed by our awareness of them; while what it doesn’t tell us is how awareness itself ought to be understood from an objective standpoint. Although Lockwood is confident a neurological explanation for the latter will one day be forthcoming, Papineau’s considerations indicate otherwise.

In view of all this, it certainly looks as if the ontology of nature which first established the bonding problem(s) needs to be challenged at a deeper level than simply denying the existence of grain. Focusing on the brain as a seat of consciousness only adds to the problem, because the complexity and ambiguity of neuroscientific evidence rules out the kind of simple identity theory Russellians need. Consequently, Goff’s initial programmatic remarks about recuperating Aristotle’s philosophy of biology do not seem quite so throwaway. Perhaps something closer to Aristotle is what is needed. A recent strain of thought (represented below by John Heil, Kathleen Wilkes, and Wesley Salmon) has questioned the existence of a structural mismatch between Aristotelian and post-Cartesian conceptions of the physical, which arises due to Cartesian doubts about the fundamentality of middle-sized objects. These philosophers raise the intriguing possibility of identifying o-properties with the categorical natures of organisms, and hence rendering the Russellian solution to the mind-body problem considerably less ad-hoc, as well as continuous with respectable biology.



### [2.3] Against Humean Metaphysics

David Lewis describes his position as one of “Humean Supervenience” (Lewis 1986, ix).<sup>23</sup> Following Hume, Lewis denies the existence of necessary connections between states of affairs; “all there is to the world is a vast mosaic of local matters of particular fact, just one little thing after another” (*ibid*). The mosaic consists of point-sized “natural intrinsic properties” related, geometrically, by their relative distance from one another. Though he regards this picture as a relevantly similar approximation to that of modern physics, Lewis is mainly interested in defending the “tenability” of his account (*ibid*, xi) against philosophical objections. Besides his physicalism, Lewis commits himself to views on causation, scientific explanation, composition, and modality which have invited criticism. Nolan (2005) lists a number of scientific advances which might force major modification or abandonment of Humean supervenience. We might find we “need fundamental properties that are indivisibly instantiated in regions, rather than just at points” (*ibid*, p.29). And quantum entanglement suggests non-spatiotemporal relationships hold among particles, and that these relationships might turn out to be the more fundamental sort out of which physical events are constructed (*ibid*, p.32-3). Lockwood gives one working example of an ontology on just these lines.

For his part, Lewis seeks to avoid commitments on QM until such a day when “it is purified of instrumentalist frivolity...doublethinking deviant logic... and ... supernatural tales about the power of the observant mind to make things jump” (Lewis 1986, xi). Elsewhere, he notes that “philosophical theories are never refuted conclusively”, but are, rather, assessed and reassessed on the basis of their relative strengths and weaknesses (Lewis 1981, p.x).

---

<sup>23</sup> To be clear, this is a metaphysical position influenced by Hume’s views on causation. A different form of naturalism, compatible with the form of scientific explanation defended by those who object to Lewis, and influenced by Hume’s response to his own sceptical doubts, is discussed in chapters 3 and 6.

Keeping Lewis's metatheoretical caveats in mind, it can hardly be expected that I will demonstrate the utter failure of Lewis's research program here. I do wish to emphasise its contingency, however, since Goff and Chalmers inherit it largely uncritically, and subsequently arrive at highly implausible views about the mind's place in nature. If a different ontology can help mitigate the bonding problems, then so much the worse for Humean supervenience. A systematic alternative to Lewis is defended by John Heil, to whom I now turn.

### [2.3.1] John Heil: Powers and Qualities

According to Lewis, laws of nature governing regularities are to be accepted as primitive, and inferred on the basis of their simplicity and range of explanatory power. This in turn lets Chalmers postulate the brute fact of brain-mind correspondence as an extra, seemingly arbitrary law of nature, whereby wherever there is informational complexity, there will be corresponding experiential complexity.<sup>24</sup> On the Humean picture of things, ordinary scientific laws are no less arbitrary. Given the contingency of cosmic law, it is assumed that there are possible worlds with completely bizarre laws (Goff considers, without endorsing, one in which certain types of blue balloons turn red when inflated<sup>25</sup>). Humeans place the burden of proof on the sceptic to show that the hypothesised laws are naturally impossible, rather than bearing it themselves. Defence of natural necessity with respect to logically contingent laws is consequently an *a priori* affair, rather than one built on appeals to empirical facts.

Heil (2015) understands Humeans to be committed to a form of modal monism in which "there are *no* metaphysical constraints on possibility" (45); only logical constraints. In

---

<sup>24</sup> See e.g. *TCM* p307.

<sup>25</sup> Goff (2014a), p85.

explaining or predicting an event, then, the Humean must appeal to its similarity to previous circumstances where the event occurred; only these can provide an inductive constraint on modality in *this* universe (there are possible worlds which differ). Rather than appealing to a vast, largely uncharted wilderness of possibility space, Heil points out that it would be more conservative to ground an object's propensities in its powers to do certain things and not others.<sup>26</sup> And he shifts the burden of proof over to philosophers who assume without argument that laws of nature are contingent (Heil 2015, p49-51).

Heil's position lends itself to scientific realism and institutionalism (deferral to scientists' expertise on ontological and modal questions). Lewis advertised his position along the same lines; deferring to scientific theories, for instance, as the best arbiter of the limits of possibility in *our* universe. However, we have already seen his reticence with regards to QM. Nolan (2005, p12) notes that this extends to other areas of enquiry; Lewis goes so far as to "rebuke the mathematicians for a foundational error" (Lewis 1993/1998, p222). And Ladyman, Ross, Spurrett and Collier (2007) list several examples of Lewisian metaphysics put to work on scientific questions of composition and causation and yielding cumbersome, implausible or misleading characterisations of events which, they argue, would have no currency in philosophy if Lewis were actually serious about his scientific institutionalism.<sup>27</sup>

The foundation of Lewis's distance from mainstream science lies, according to Heil, in a purely philosophical mistake, rather than (as Ladyman *et al* argue) a failure to engage with science at a higher grade than A-level chemistry and Newtonian physics (Ladyman *et al*, p24). The mistake is Lewis's absolute separation of powers and qualities, which roughly corresponds to the separation of extrinsic and intrinsic properties. Humeans postulate

---

<sup>26</sup> This is a throwback to Aristotle (see Shields 2013, p68-9), who denied the existence of sheer possibility, that is, modality that makes no reference to a thing's actual potentials for modification. See chapter 3, below.

<sup>27</sup> Wilkes (1988) likewise argues (p.42) that the ambiguities of speed and location in the quantum-mechanical world picture are of immediate relevance to what, if anything, it could mean for there to be an exact duplicate of our world, and thus for the idea of a world in which 'all the facts' are known – consider the importance of Heisenberg's *uncertainty* principle.

properties that are non-causal – those on which causal properties supervene – which consequently cannot be known by science. Hence they allow for possible worlds whose relational properties are identical to our own, but which differ with regards to the “quiddities” underlying them.<sup>28</sup> The Humean picture loses sight of essentialism’s utility as an empirical hypothesis: the ‘essence’ of water is not a second layer of reality without causal powers of its own.<sup>29</sup> That essences themselves can have effects on us is crucial to Kripke’s story of how we come to refer to them in the first place. By contrast, Goff’s twist on this view motivates him to doubt that science even has a grasp of relational properties. In the co-authored paper, “What’s Wrong With Strong Necessities?”, Goff writes that

it is not obvious that to conceive of water as H<sub>2</sub>O is to have a transparent conception of water: it may be that our chemical concepts refer by description. In this case, we can imagine replacing the concept of water with a concept of H<sub>2</sub>O that conceives of it in terms of its intrinsic nature. (Goff and Papineau 2013/4, p754)

Elsewhere he has proposed that “even empirical investigation won’t reveal the essence of wine, as observation reveals only the extrinsic features of material things” (Goff 2014b, n. 17, p19). At this point the relation between Lewisian metaphysics and scientific realism seems to have been truly severed, allowing Goff to speculate freely about the non-mathematico-spatial relations upon which o-properties must supervene. But the motive for doing so lies in an impoverished ontology which separates essences from empirically measurable properties, and foregoes powerful properties altogether.

---

<sup>28</sup> Esfeld (2010, p128-32) argues this point at some length in his critique of Humean metaphysics.

<sup>29</sup> Esfeld (2017). See p219-222 on how matter came to figure as the bearer of, and explanation for, the causal effects studied by early scientists (chapter 3, below, emphasises their distance from medieval conceptions of matter).

Heil (2015) names his alternative “Aristotelian supervenience”, although in the past he has denied the picture normally implied by supervenience relations, whereby reality is separated into “levels”. Heil (2003b), for instance, targets the foundations of the Canberra Plan, whereby realism about anything “is a matter of something’s possessing a particular property answering to the predicate” (208) by which we refer to it. Heil argues that this leads to the reification of dispositions or standing properties, which are multiply-realizable, into distinct ‘levels’, thereby granting them determinacy in spite of their indeterminate or ambiguous range of realizers. This then generates the further problem of causal overdetermination or epiphenomenalism, but the founding error, Heil believes, lies in “a correspondence principle that licenses our ‘reading off’ features of the world from features of our language” (2003b, p218). But it is unlikely that there could be a distinct sentence for every possible feature of the world, and, hence, it is not clear how those features could *necessarily* entail the truth or falsehood of some sentence, as required by logical supervenience. On the contrary, “concepts, and words used to express these, are in most cases satisfied by endless *similar* things; and similarity grades off imperceptibly into dissimilarity” (*ibid*, p219).

Boundaries between things are not just arbitrary, however, nor do they have to be thought of as intrinsically vague. According to Heil, individuation is marked not by disjunctions of composition (as in Sider’s analysis of tables), but by powers. However, Heil is also keen to distance himself from philosophers, such as Ladyman *et al* (2007), who view the world as a kind of network of relations or forces acting directly on one another, rather than on passive substances. Such a worldview yields no answer to the question of what powers or dispositions consist in, save for an infinite regress: As comprise the power to cause Bs, Bs comprise the power to cause Cs, etc (Heil 2004b, p237). *Prima facie*, then, we have a motive for introducing objects into our ontology in order to serve as the bearers and subjects of

(powerful) properties, and as the referents to which our concepts are predicated (since analysing things directly into bundles of properties seemed to be a dead end). Conversely, possession of certain capacities to act (and, in the case of life, ongoing activities such as seeking nourishment) was, for Aristotle, the criterion for individuating forms, i.e., objects. As Heil (2004a) puts it, “a property is nothing more than an object’s being a particular way” (p.193). So, perhaps an object can just be whatever it is appropriate to ground the analysis of some properties in.<sup>30</sup>

The centrepiece of Heil’s ontology, however, comes from his claim that powers are inseparable from qualities, since it is the latter which provide an intuitive answer to what the presence of certain dispositions could amount to when they are not being exercised. While this once again harks back to Aristotle, specifically the claim that “form generates quality and matter generates quantity” (Shields 2013, p.199), Heil also wishes to reassess John Locke’s distinction between the primary and secondary qualities of material objects. Heil (2004b, p240-1) considers an argument from Armstrong (1961, ch.15) to the effect that a universe containing entities composed solely of primary qualities (shape, size, position, duration, movability, divisibility, and solidity) would be indistinguishable from empty space, since empty spatial regions can possess all of these qualities (he points out solidity can be reduced to impenetrability, and obviously one spatial region cannot penetrate another). Armstrong contrasts these to the secondary qualities of colour, sound, taste, and smell, which between them ground the reality of objects distinct from the space they inhabit; but he thinks these properties will have nothing to do with the powerful primary qualities. By contrast, Heil thinks primary and secondary qualities can be brought much closer together, arguing that the latter are simply primary qualities “distinguished by...their effects on conscious observers” (2004b, p244). Hence, “being spherical is a manifest quality of a

---

<sup>30</sup> Heil’s ontology turns out to be much less liberal than this, as I will discuss in [3.4.3]. But as we will see shortly, it would not help solve the bonding problem even if it were.

baseball”, and this quality is also a power to roll (ibid 243). The quality itself can be considered mind-independent: being sweet explains why a cake manifests to us as tasting sweet. The separation of powers from qualities is an act of imagination which we impose on objects of experience. However, the individuation of objects themselves is not, since the “power net” of which they form part is independent of us; e.g. the presence of oxygen in disposing a match to light when struck (2004a, p202). The network can even include socially constructed objects, such as statues, whose causal history ensures they are to be understood as more than lucky accidents (ibid, p205).

Heil’s place for qualities bears much resemblance to the Russellian place for qualia as the realizers of extrinsic properties; he argues qualities realize dispositions (2004a, p198). In fact, Heil compares his views with Lockwood’s (Heil 2004a, p242), and follows panpsychists in attributing a primitive quality to fundamental particles: “Perhaps, insofar as an electron can be thought of as occupying a fuzzy region of space, an electron could be thought of as having something rather like a shape” (2004b p244-5).

### **[2.3.2] Wesley Salmon on Scientific Explanation**

Heil’s alternative to Humean metaphysics would not be complete without reference to the philosopher of science, Wesley Salmon. Salmon’s portrayal of scientific understanding differs substantially from the so-called deductive-nomological (D-N) school of thought associated with Carnap (1936; 1937) and Hempel (1962), as well as Popper’s falsificationist account (Popper 1935/1959). These latter focused on laws, or inviolable generalisations, which could apply to a great range of phenomena; or failing that, generalisations with a high degree of probabilistic certainty, as in Hempel’s inductive-statistical (I-S) amendment to D-N. The I-S and D-N models both liken scientific explanation in general on the success of lawlike

explanation in fundamental physics (Bechtel 2007, p3-6). Humean metaphysics inherited this picture, not only because it tends to identify the ultimate essences or quiddities with the smallest building blocks of nature, but also because of the shared presupposition of the powers of analytic, a priori reasoning. Proponents of the deductive-nomological philosophy of science may not be scientists themselves, but they can clarify the foundations of scientific explanation, if these are construed as a matter of deductive logic or falsifiable claims made on the basis of sensory experience. This gave philosophers authority in describing the context of justification behind a scientific discovery, while ignoring the context of the discovery itself, which could include any number of apparently contingent and merely psychological factors (Bechtel 2007, p34). By situating the bulk of scientific rationality at the a priori end of knowledge, these philosophers were able to present themselves as equal partners to scientists and, via the logical analysis of concepts (e.g. 'consciousness'), could advance claims about the necessary and sufficient conditions of explanation that compete with scientists on their own turf. Chalmers' declaration that a science of consciousness has not yet even begun – because it has failed to analyse phenomenal qualities – is merely the latest example of this trend.

By contrast, Salmon's account portrays science as a much more piecemeal and ad-hoc process. Rather than seeking contingent (but ideally exceptionless) laws governing highly regular occurrences, for Salmon a scientific explanation can encompass any event, even one which only happened once (he points out that this is not possible on Hempel's I-S model). To see the difference, Bradie (1996) emphasises Salmon's appeal to the notion of a Laplacian demon, with complete descriptive knowledge of all particles and their velocities throughout the universe. The demon



would have no understanding of nor any need to appeal to underlying causal mechanisms. Such a demon, however, would not be in a state of perplexity about anything and so the "why?" questions which arise from states of incomplete knowledge and which generate explanatory accounts would not occur to the demon (Bradie, 1996, p315-6)

The D-N model of science aspires to such a demon's comprehensive knowledge, where knowledge is equated with a predictive argument – if X, then Y – whose terms should, again ideally, consist of both knowledge of the total state of the universe at time 0, and an inference as to the total state of the universe at any time thereafter. The inference from X to Y ideally requires having 'all the facts' because "[i]nference – whether deductive or inductive, demands a requirement of total evidence – a requirement that *all* relevant evidence be mentioned in the premises" (Salmon, 1977, p151). Whereas "an *assembly of facts statistically relevant* to the explanandum" (Salmon 1984, p. 45) does not. As Maria Galavotti puts it, "an event is explained by showing what factors are statistically relevant to its occurrence", and "rul[ing] out all irrelevant ones" (Galavotti 1999, p40). The Ideal of explanation is "homogeneity" – where, all things being equal, *only* those circumstances described in the explanation are relevant to the occurrence. Salmon

regards statistical explanation as the general case, and explanation making use of universal generalizations (deductive-nomological explanation) as a special case. Embracing a probabilistic perspective compels him to take the notion of relevance seriously, for probability judgments depend in a crucial way on relevant information." (Galavotti 1999, p39)

The focus on circumscribed explanations over comprehensiveness feeds into Salmon's account of causation, which aims to "put the 'cause' in 'because'" (Salmon 1977, p160). With

this Salmon “does not commit himself to metaphysical claims or anti-Humean attitudes” (Galavotti 1999, p43). Phil Dowe (1992) points out that “one of Salmon's broad objectives is that causality be understood as "contingent"... Salmon does *not* regard the task as one of providing a conceptual analysis applicable to all logically possible worlds” (p199). Instead, explanation describes a “causal process”, composed of “physical properties” that are “taken as a spatio-temporal continuous entity” (Galavotti 1999, p42-3); where processes intersect it will be possible to account for what information, or ‘mark’ they bear on one another, which forms the criterion for the transmission of causal influence. This “essentially geometrical model” (*ibid*, p45) of causal relations can be compared to a map or a telephone network (*ibid*); it is highly favourable to a static or ‘block’ theory of time. It also bears close comparison to Aristotle’s approach to time and causation, in which causes are contemporaneous with their effects (Shields 2013, p77-8).

Whereas the structure of the cosmos over the course of its history is for Humeans as contingent as its sheer existence, for Salmon explanation seeks to account for the network of causal influences as an internally consistent whole; only the reasons for its existence remain open to sceptical doubt. Indeed, “causal explanation is wholly compatible with anti-realism” (*ibid* p51), particularly pragmatism, since “information about mechanisms could quite naturally be related to manipulability” (*ibid*). Nevertheless, Salmon’s account is also attractive to realists, as it draws a close connection between the epistemic activities of scientists and the “ontic” properties of their objects of enquiry, since to provide an explanation is to point to a mind-independent process or mechanism of a certain kind.

Crucially for current purposes, the explanatory strategies employed by scientists in the field are therefore relevant to philosophers’ ontologies, as Bechtel (2008) argues on behalf of biologists:

...in many parts of biology one seems to look in vain for what philosophy has commonly taken to be the principal explanatory tool of science, that is, laws. The few statements that have been called laws in biology, such as Mendel's laws, have often turned out to be incorrect or at best only approximately correct ... that does not mean that biologists and psychologists are not developing explanations. If one investigates what biologists and psychologists seek and treat as sufficient for explanation, it often turns out to be mechanisms, not laws. (Bechtel 2008, p10)

Ernst Mayr, a biologist responsible for the 'grand synthesis' of molecular biology and evolutionary theory, and later a philosopher of science, lists "[i]ndividual differences, communication systems, stored information, properties of the macromolecules, [and] interactions in ecosystems"<sup>31</sup> among the objects of biological study which do not admit of precise analysis into necessary and sufficient physical conditions. He claims that biologists are explaining *types* of pattern invisible to the lower levels of description: "species, classification, ecosystems, communicatory behaviour, regulation, and just about every other biological process deals with relational properties"<sup>32</sup>. Making sense of these relations presupposes real objects bearing the properties in question. And while Mayr otherwise associates essentialism with 17<sup>th</sup> Century metaphysics, he does have kind words to say for Aristotle's *eidos*, or spirit, and for the Leibnizian concept of a monadic (absolutely individual) property.<sup>33</sup>

Mayr's case for biological essences appeals to the form-bestowing capacities of genetic material, which has proved controversial. It has been argued that Mayr reifies what is at best the metaphor of DNA as a program or instruction manual for the organism (Vinci

---

<sup>31</sup> Mayr (1982), p55.

<sup>32</sup> *Ibid* p56.

<sup>33</sup> "Aristotle's *eidos* (even though considered immaterial, because invisible) was conceptually virtually identical with the ontogenetic program of the developmental physiologist." (*Ibid*).

and Robert 2005, p204), and that in any case this notion of form imposed on the organism from without is a misreading of Aristotle (*ibid* p205; see also Shields 2013, p98). Some other scientific criteria for the functional unity of an organism could be supplied, however, so these worries are not a knock-down case against the sort of essentialism which could find a place for o-properties in nature.

However, we should remember Salmon's philosophy of science goes hand-in-hand with naturalism, where philosophy is continuous with scientific explanation. While the question 'why did X occur?' does not a priori rule out appeals to the kinds of intrinsic properties Chalmers and Goff defend – e.g., in accounting for the truth of phenomenal judgements – the trajectory taken by scientific psychology suggests that such explanations are far-fetched. Kathleen Wilkes argues that the only kind of biological essentialism that can account for both the unity of personhood and the disunity of phenomenal judgements will have to dispense with consciousness as a natural kind altogether. If she is right, then powers-based metaphysics faces a structural mismatch problem which prevents non-reductionist biology from reconciling with Russellian monism, including the promising approach recommended by Heil.

#### **[2.4] Wilkes on Personal Identity**

Drawing on evidence amassed from psychology, Wilkes argues that while science is in a position to discover natural kinds, 'consciousness' will not be among them. She defends a kind of eliminativism about the mind, although she would reject this label of her position, since unlike full-blown eliminativists such as the Churchlands she thinks of our common-

sense psychology as a useful fiction which is unlikely to ever be entirely displaced.<sup>34</sup>

Nevertheless, her argument against realism about consciousness is similar to Paul Churchland's argument against beliefs and desires in Churchland (1981). The way Wilkes sees it, insofar as consciousness presents itself as a distinctive problem to scientists, it manifests as a heterogeneous group of phenomena which could more profitably be classed separately; on the other hand, there are reasons, drawn from case studies, for thinking that there is no clear distinction between conscious and unconscious states anyway. She would therefore support Papineau's scepticism about the determinacy of consciousness, discussed in [2.2.2]. So we are yet again faced with the bonding problem: on the one hand, the problem of what unites various distinct, objective phenomena of interest to psychologists under the heading 'conscious'; on the other, a variation on the decomposition problem, where distinguishing the conscious subject from an undifferentiated field was the major problem facing Lockwood's many-worlds-sized qualities, and Goff's cosmos-sized psyche. Narrowing down the latter to a body-sized field does not get us far enough if most of what characterises the body's activities is unconscious.

Among the many areas of interest in psychology is accounting for the difference between sleep and wakefulness. But this also means accounting for dreaming, sleepwalking, hypnosis, fugues, and epileptic automatism, where malfunctions in short- and long-term memory, among other issues which incline us to say somebody was 'not themselves' at that moment, raise the question of whether there was awareness at all. Subjects who undergo these altered states will often fail to recall them, but behavioural evidence – a mix of automatism and purposefulness – leaves us none the wiser either. The mystery deepens when more unnatural states such as blindsight are taken into account. Patients afflicted with this condition, the result of damage to visual processing regions of the brain, will deny they

---

<sup>34</sup> Wilkes (1988), p13. This seems to have been a recent change of heart at the time, having argued in favour of eliminating consciousness from our vocabulary in "Is Consciousness Important?" (1984).

have any visual experience, and yet demonstrate behaviour consistent with some limited vision at an 'instinctual' level. Wilkes takes this as evidence that conscious sensations and consciousness *of* need to be distinguished,<sup>35</sup> in the sense that folk psychology seems to attribute some form of awareness of their environment to blindsight patients (Wilkes's willingness to take the intuitions of the folk at face value, in spite of their vagueness and inconsistency, is to some extent parasitic on her account of personal identity. The issue is further discussed below).

In any case, she says, it will be up to science rather than philosophy or common sense to decide, on the basis of all this evidence, whether to be a sensation is to be perceived (1984, p226). Wilkes denies there is common ground between sensations occurring in the body, and perceptual states relating to the environment, since she denies there is any sense to be made of qualia (see below). She argues that the category of the mental is heterogeneous with respect to intentional states, such as thoughts, beliefs, and self-consciousness; any of these could occur in the absence of the others, and most could be construed (again, by folk psychology) as taking place unconsciously. But divided attention, weakness of will and self-deception seem to call first-person authority into question even when a subject is unambiguously conscious (1988, p102). Wilkes concludes that, so far as science is concerned,

classification in terms of consciousness or its absence is simply too crude to cope with all this diversity; surely profitable research into these and other phenomena will require a theoretical classification determined along different, and probably more fine-grained, principles." (1984, p231)

---

<sup>35</sup>Wilkes (1984), p226.

Echoing Churchland's focus on neuroscientific case studies, Wilkes gives the example (1988, p158-160) of pure alexia, a condition which leaves individuals able to write but not to read. Folk psychology and non-naturalistic philosophy are at a loss to explain the loss of one ability and not the other; only neuroscience can account for it, in a 'non-rational' way. She concludes that "non-rational [i.e., mechanistic] explanation has *supplanted* the rational/intentional model (just as meteorological explanations of thunderstorms have replaced rational/Intentional anthropomorphic explanations via angry gods)" (1988, p160). Unlike Churchland, however, Wilkes denies that this carries across the board to all familiar folk psychological terms; the "wants, reasons, hopes, fears, decisions, calculations, interests, beliefs" (1988, p157) which Churchland would have us do away with altogether. This is because she thinks these terms can be understood instrumentally, and have only one ontological commitment, the unity of a person (1988, p23), which is not usually in doubt.

The most striking example of conscious disunity discussed in philosophy is the case of split brain patients, whose brain hemispheres have been prevented from communicating with one another in a last-ditch surgery intended to prevent epileptic seizures. Even Thomas Nagel (1979) worries that this example flummoxes our usual notions of personhood; once again only neuroscience seems to even have the vocabulary to describe what is going on. But Wilkes argues that there is a deep behavioural unity to split brain patients, and that the sheer capacity to be 'in two minds' does not immediately call this into doubt;<sup>36</sup> "The disunity holds against a background of 99 per cent unity" (1988, p156-7). Even in carefully designed experiments, forcing disunified behaviour is hard, and patients dislike it, taking whatever steps they can to minimise the disharmony. Wilkes points out that the brains are never entirely split, being divided at the the cortex (*ibid*, p37-8), so the split is "only" at a conscious level (*ibid* p151), and even then is no more or less mysterious than divided attention,

---

<sup>36</sup> She has a very different response to sci-fi scenarios where people split and fuse indiscriminately. See next chapter.

weakness of will, self-deception, etc (*ibid* p146-7). Furthermore, brain damage caused by previous epileptic seizures might explain some of the patients' behavioural discontinuity without needing to postulate two autonomous and discretely functioning hemispheres or 'points of view' (*ibid* p151-3).

Preserving most of common sense psychology in light of scientific evidence, then, is bought at the price of jettisoning consciousness from 'serious' ontology: "The mind contains, and must be recognized to contain, depths to which such an 'I' has no access" (1988, p221); "consciousness [is] no more than the bare tip of an enormous and largely unknown iceberg" (*ibid* 82); [m]uch more of our routine sensory experience is like this than we perhaps care to realise; it escapes our attention, maybe, because when we consider and concentrate upon our sensory experience, we alter the phenomenon we sought to describe." (1984, p227) Wilkes strongly suggests that the apparent importance of consciousness to our self-understanding is conjured up by a kind of intellectual self-reflection, the popularisation of which as a source of philosophical insight was, she argues, started by Descartes, for whom

Sensation ... was .... redescribed as a kind of thinking. But thinking is hard to think about without the metaphor of sight—of *ordinary* sight. The viciousness of this circle is impossible to exaggerate. Ordinary sight has objects to look at. Hence, it seems, thought must have objects, and objects analogous to public objects of sight in the real world, mental *entities*: 'ideas', 'impressions', 'sense data', 'qualia'; these are soon to be construed as 'mental atoms' by parallel with Newtonian atoms. Thus we get the second domain of the mental, and a double ontology. That foists on us the dualism of the inner and outer realms; and we now cannot understand either thinking or perception." (1988, p219)

Wilkes rejects this model of the mind as an inner light. She argues (1984, p229-30) that experience is largely of public objects rather than inner sensations, and, in any case, that



sensations do not need to be reflected upon in order to exist. Insofar as the phenomenon of consciousness, as opposed to thought or sensation, exists at all, it is either an illusion caused by reflection, or it is predominantly a pre-reflective phenomenon with objective properties which reflection could misrepresent. Neither possibility is friendly to anti-physicalist arguments, where the appearance of the mental is its reality, and both have recently been proposed by Keith Frankish as 'illusionist' physicalist solutions to the hard problem.<sup>37</sup>

Frankish identifies illusionism as the "claim that introspection delivers a partial, distorted of view of our experiences, misrepresenting complex physical features as simple phenomenal ones ... We can introspectively recognize these states when they occur in us, but introspection doesn't represent all their detail. Rather, it bundles it all together, representing it as a simple, intrinsic phenomenal feel" (Frankish 2016, p18). He therefore suggests a cognitive explanation of phenomenal bonding is all we need, and that appeals to fundamental ontology, exotic physics, or biology are unnecessary.

In a further parallel with Wilkes' historicism, Frankish also argues that qualia are not a well-defined basis for arguing about the mind because people can disagree over whether they are features of sense data or external objects. But construed as properties of mind-independent objects, qualia could not be "directly introspectable".<sup>38</sup> In line with Wilkes (who herself seems to favour some variation on the idea that consciousness discloses objects in the external world rather than a Cartesian theatre), Frankish suggests that the fact that we can reconceive experience this radically should inspire scepticism about the existence of scheme-neutral contents of consciousness. Perhaps, he says, all there are are judgements about what we experience (*ibid*). I return to the issue of historicism about concepts of consciousness in chapter 3, and to illusionism in chapter 5.

---

<sup>37</sup>"Illusionism as a Theory of Consciousness", in Frankish, K., (ed.), *Illusionism* (2016).

<sup>38</sup> Frankish (2012), p670-671.

### [2.4.1] Panexperientialism Again

All the same, the parallels with panexperientialism (or rather, with the paradoxes which result from it) should also be clear as well. The unconscious unity of the body could certainly be identified with Whitehead's 'experience', while the appearance of consciousness as a separate, and separable, phenomenon (the tip of the iceberg, as Wilkes put it) from that of the body could be due to the reflective intellect. Maybe Wilkes's criteria of personhood individuate a person's protoconscious field. Wilkes considers this herself:

...qualia, if they are to be postulated at all, must equally be postulated at a non-conscious level; the role they are ascribed is that of explaining how we distinguish objects on the basis of their colour (or other sensory) qualities, but subliminal perception too involves colour discrimination (1984, 234)

Lockwood discusses Wilkes on unconscious unity in order to reject her solution as "neither here nor there" as regards the problem of understanding conscious unity (1990, p83-4), but his dismissal is due to reading her as a straightforward eliminativist. Like Wilkes, he still takes "familiar cases of absent-mindedness, divided attention and so forth as relevantly similar to the split-brain cases" (*ibid*), and goes along with Wilkes' attribution of a lower level unity of personhood at the base of the brain where, he points out, emotional responses are processed (*ibid* p89), just as feelings were for Whitehead more fundamental than thoughts. Moreover, experiences can be objectified: Lockwood borrows from the impersonal language of set theory (*ibid* p88-91) to describe the differing levels of relations and differentiations at work in split brains and ignores the question of which set of experiences I would be if my

corpus callosum were cut.<sup>39</sup> He confidently states that “all that is required to imagine a given state of awareness of a commissurotomy patient is to imagine each of his current, overlapping phenomenal perspectives in turn” (*ibid* p92), when it is surely the difficulty of imagining each simultaneously that poses a philosophical problem. Further underscoring that he has missed the point, Lockwood writes that

What is accessible to any act of introspection will invariably be the contents of a single phenomenal perspective. For that reason, commissurotomy patients aren't going to *feel* disunified either. The issue is just *how many* overall unified-feeling states of awareness — that is to say, phenomenal perspectives or maximal experiences - there are at any given point in the biography of such patients, or of ourselves. And that is a question that sheer looking within oneself can do nothing to answer. (*ibid*)

What unites panexperientialism and physicalism is the claim that experiences are real, objective things which exist in some way ‘mind-independently’. This is straightforward in the case of physicalism and naturalist epistemology, where the identification of sensations with the objects of psychophysics makes room for revisionism, as Wilkes’s claim that psychology has dispensed with consciousness illustrates. But Chalmers likewise separates qualia from phenomenal judgements about them; the only difference is that he denies science is in a position to study their properties. The logical conclusion is something akin to Lockwood’s subject-deflationism, where, as quoted above, introspection is no longer treated as a reliable guide to phenomenology. (As Frankish puts it, for self-scanning to be mechanically possible implies the possibility of error<sup>40</sup>). But if we cannot know if we are one or many from looking within ourselves, then we lack the authority to judge whether consciousness is an all-or-

---

<sup>39</sup> His position is therefore close to Lichtenberg’s, for whom ‘there is’ experience, without reference to a bearer or subject of experience.

<sup>40</sup> (2016), p30.

nothing property, or whether constitutive panpsychism is true; and in that case, as Goff has argued, we might as well give up and become physicalists instead.<sup>41</sup>

Goff's solutions faced major problems with mental causation, due to the question of how o-properties relate to the physical. So perhaps a reimagining of causation along Russellian monist lines would help clear things up, and this is exactly what Gregg Rosenberg has sought to supply. Rosenberg synthesises and builds on all the positions put forward in this chapter, and uses them to finally explain what Goff thought was impossible: describing the non-physico-mathematical relations which bind consciousness, which he calls "the mesh".

## **[2.5] Rosenberg's Ontology**

So far, it has begun to look as if any form of panpsychism faces serious difficulties, with holistic ontologies, which seek to avoid the grain and composition problems, falling prey to decomposition and structural mismatch problems instead, and resulting in positions no more plausible than constitutive panspsychism or physicalism. The question remains as to what, if anything, in the physical world could count, from the third person perspective, as the instantiation of a subject.

Crucially, Rosenberg's solution builds on the negative assessment of Humean metaphysics discussed in previous sections, which called into question Chalmers' basis for posing the hard problem in terms of the failure of the mental to logically supervene on a base of particles banging into each other (as Ladyman *et al* 2007 put it, p.3). He sides with

---

<sup>41</sup> Goff (2009a), p129-135.

Chalmers' "more fundamental" argument<sup>42</sup> against physicalism by appeal to the explanatory gap between physical and phenomenal properties, "making a direct argument from what Chalmers calls the "absence of analysis" and not an indirect argument from conceivability or new knowledge (Rosenberg, 2004, p26-7). As will become clear, Rosenberg's ontology extends this in a way quite congenial to Aristotle, for whom there could be no further analysis of substantial forms.

Rosenberg agrees with Heil and Salmon that a robust form of causal explanation must go beyond this empiricist picture of the universe as a container of disconnected objects, and of science as simply a catalogue of the objects' constant conjunctions with one another.<sup>43</sup>

Causal interactions imply partitions: they divide the world into different, mutually influencing spaces, and do so at many levels of nature. We might discover, by looking at causality and causal interaction, that causal interactions have certain important aspects which distinguish natural individuals. For panpsychists, these individuals would be attractive candidates as the supporting ontology of experiencing subjects." (Rosenberg, 2015, p227)

Rosenberg argues, however, that construing explanation in terms of causal powers doesn't go far enough. He calls the latter an example of 'causal responsibility'-based explanation, in the sense of X bringing about Y. Thinking of explanation in these terms renders negative facts, such as loss of life due to lack of food, difficult to understand; "facts about absence

---

<sup>42</sup> Chalmers (2010), p.xv.

<sup>43</sup> "...a core, unargued assumption in much of classical metaphysics is that things at lower levels of nature are already, in and of themselves, in determinate states, just as we find things at higher levels in ordinary experience. It is also assumed that things at higher levels of nature *must* inherit their determinate states from the determinate states of the lower-level things which compose them. In philosophical parlance, classical metaphysics assumes that determinate macro-level states of the world *must* strongly supervene on independently determinate micro-level states of the world. This statement is a kind of *micro-determination thesis*, or *mic-d* for short." (Rosenberg, 2015, p231)

require appealing to intentional objects such as universal “That’s all” facts” (Rosenberg, 2004, p147). This picture of causation is still too loaded with folk metaphysics for Rosenberg: “the ideas that causal relations are asymmetric, that they exist only forward in time, that they are only local in space, perhaps that they involve events, and that [causal responsibility] is a two-place relation.” (Rosenberg 2004, p148)

All of which makes sense if the starting point for causal explanation is “why did something happen?” (*ibid* 150), leading one back to the ancient question of why anything exists in the first place (*ibid*); as [2.3.2] noted, this was left outstanding by Salmon’s account of causation. But Rosenberg points out that these questions can just as well be reversed: we might wonder why only some things happen rather than others, and why not every possible thing seems to exist. He proposes a major modification of Salmon’s account. In place of causal responsibility, Rosenberg puts forward causal *constraint*: what restricts the actual to certain possibilities over others, including what causal powers (what he calls effective properties) are instantiated. “[F]undamental physics of our universe will be the science that *at least* discloses to us the effective dispositions of the fundamental individuals of our universe, assuming such individuals exist. However, none of this implies that physics will yield a complete account of the world’s causal structure.” (Rosenberg 2004, p153)

Describing effectivity is as far as powers metaphysics goes (*ibid* p298). In contrast, Rosenberg adds receptivity: the disposition to be e/affected. This is not a property with any power to affect others (he calls it a paradoxical “passive power”, p154), and so would seem to be surplus to the scientist’s explanatory needs. But it is also surely a conceptual requirement for causal power to have objects upon which they are actually able to act. In other words, there must be Aristotelian substances.

To be receptive of causal determination one must first be a determinate individual, according to Rosenberg, and these “could exist at any level of nature ... so there would be no

special pride of place given to microphysical individuals” (Rosenberg 2004, p157). But what it is that causal powers affect is empirically underdetermined, being subject to interest-relative assumptions about what to class as an individual, which in turn rests on questions of causal *responsibility*, not constraint. As Salmon himself puts it, “pragmatic considerations determine whether a given 'process' is to be regarded as a single process or a complex network of processes and interactions.... It all depends upon the domain of science and the nature of the question under investigation” (Salmon 1994, p309). Rosenberg clarifies that “physical theory leaves out information about receptive connectivity [also termed causal nexii or subjects] and the intrinsic carriers [qualia]” (Rosenberg 2004, p9); “nothing in physical theory will give logically conclusive reasons to assert the existence of one structure instead of one of the others”, he writes (*ibid* p223). Hence it is unsurprising that Mayr’s attempt to furnish scientific criteria for biological essences turned out to be underdetermined and metaphorical. Rosenberg is against simply identifying ontology with scientific explanations: “societies may appear as individuals within sociology, and galaxies may appear as individuals within astronomy, but it does not follow that they are *natural individuals*. The *natural individuals* ... are individuals in virtue of the fact that they have a special, unitary causal nature” (*ibid*, p178-9). Jennifer McKittrick (2006) glosses Rosenberg as saying that “science provides the nomic mosaic – the simplest description of the lawful ways effective properties instantiate through space-time” (6), and Rosenberg argues at length that this ‘mosaic’ of constant conjunctions can be filled out with many different theories, which postulate quite different individuals (Rosenberg 2004, 219-228).

Rosenberg points out that “the bound individuals [i.e., particles] within the nexus [i.e., person] do not depend on the existence of the nexus. From the perspective of a third-party observer, the high-level receptivity of the individual would be far less striking than the hierarchy of individuals it directly and indirectly binds.” (*ibid*, p254-5) His reasons for

thinking that lower-level states should *appear* more determinate than persons are complex, but essentially amount to a variation of Lockwood's views. For Lockwood the determinacy of particles relative to the ambiguous smudge of both the perceptual field (which is invisible to third-person science) and the multiverse as a whole (which was likewise invisible to science prior to QM) was down to the "arbitrary selectivity" of our perspectives on mind-independent objects; somehow the determinacy of *this* subject and *those* particular physical states are mutually compatible and presupposing. In contrast to Lockwood's mysterianism, however, Rosenberg thinks he has accounted for this selectivity of perspective, at least to some extent, since organisms come out in his ontology as especially determinate entities.

This explains why consciousness need not show up as a natural kind for psychologists seeking to explain and predict human activity, as Wilkes argued was the case. Receptive properties of causation are metaphysically necessary rather than empirically given. For this reason, they can account for the underdetermination of mind-body identity claims; the possibility of multiple mutually exclusive answers as to what a subject was conscious of when, and even to how many subjects are instantiated in one body, all of which Wilkes thought made nonsense of the mind-body problem. Ultimately, Rosenberg argues, the correct answer to what the receptive and effective carrier, or subject, of causal powers in organisms is can only be known by *being* that organism (or rather, the natural individual by which the organism's structures and powers are constrained): "within a cognitively structured individual, binding can provide a kind of knowledge by acquaintance; and one can deduce the privacy of an individual's experience from the privacy of its receptive field."

(Rosenberg 2004, p270)

Strictly speaking, Rosenberg's ontology comprises both "effective intrinsic carriers" of causal powers, represented by qualia, and "receptive carriers", or subjects; "neutral essences with a kind of inherent openness to their nature that can be filled by determinable



properties.... In a physical state of sensory isolation.... meditative experiencers consistently report achieving a mental state that they identify as “pure” awareness in which consciousness is perceived as possessing a kind of contentless openness.” (242) (I say more about Rosenberg’s deference to mysticism here in the next chapter.) The relation between these two types of carriers is likened to the relationship between cans of Coke (effective carriers) and the plastic rings (receptive carriers) binding them into a six pack (168). As the metaphor suggests, the receptive subject is Rosenberg’s solution to the phenomenal bonding problem (see Rosenberg, 2015, p239). His account differs slightly from Goff’s, where o-subjects or ghosts are themselves simply ‘big’ quale, composed of many phenomenal qualities, rather than existing as the matter formed by powers/qualities (and Rosenberg will deny that any of the mental entities he postulates can exist separately from the causal mesh).

Rosenberg rejects powers as the foundation of causal explanation due to the ambiguity of causal responsibility, and, consequently, its indifference to norms and mental causation (*ibid* 147). Wilkes bears this out, as she is thoroughly sceptical about our acting for clearly defined reasons; she thinks it is pointless to attribute complex motives to people where there is any room for doubt, “For we rarely pursue far the implications of our explicitly held beliefs; were we to do so, then it is plausible to suppose that all sorts of hidden inconsistencies and contradictions would come to light” (1988, p102-3). So she should be quite accommodating to the meaning-scepticism resulting from Quine’s indeterminacy of translation,<sup>44</sup> which in turn is the ultimate result of seeking purely empirical sources for explaining a subject’s words and deeds.

Rather than seeking to construct a coherent context, in which determinate thought and experience could exist, out of the overlapping tangle of material processes known to

---

<sup>44</sup> *Word and Object*, 1960, ch.2.

science, Rosenberg argues the subject itself provides the context in which lower level things are constrained (“bound”). As McKittrick phrases it, “The sub-personal parts of a person are only indeterminate and incomplete insofar as they are considered independently from the causal nexus in which they figure” (2006, p16), and so the individual which they compose provides the answer to *why* they are organised into an individual. Referring to the ambiguities of QM, Rosenberg writes that the world is “inherently *indeterminate at the lowest level but made determinate by the presence of higher-level individuals...* The role of higher-level individuals is to act more like final causes” (Rosenberg 2015, p236-7); the determinacy of lower levels is, as it were, *post-hoc*.<sup>45</sup>

It goes without saying that Rosenberg is aiming for a holistic solution to the phenomenal bonding problem, but a more interesting consequence of his ontology is a solution to the opposite, decomposition problem as well. He draws on Keith Turausky’s idea “that the base quality of the Russellian universe is like white noise or white light”<sup>46</sup>:

White noise is a relatively simple quality that masks a complex structure, which superposes all possible sounds. It is like playing twenty thousand tones all at the same time, which manifests as one single wave form ... containing within it the superposition of all possible qualities. Imagine at the basis of the universe’s quality space there is a ‘fundamental tone’...  
(Rosenberg 2017, p172)

Whereas for Lockwood the whole blur of the multiverse is a determinate entity, for Rosenberg mind-independent reality is a much more nebulous sort of thing, a “possibility space” (Rosenberg, 2004, p205) from which determinate entities are subtracted. He classes

---

<sup>45</sup> Goff (2017a, p151) characterises Rosenberg as defending layered monism, which in this context does not imply dualism, and seems to be a reasonable way of thinking about what Rosenberg is saying.

<sup>46</sup> Rosenberg (2017). Turausky presented this idea “in the summer of 2012, at a philosophical workshop on the combination problem organized by David Chalmers” (Rosenberg 2017, p172). Goff (2017a, ch.8) likewise writes approvingly of Turausky’s thesis.

this as a form of “abstract modal realism” (*ibid* p206), because so-called “level zero” individuals populating this realm exist outside of any determinate context. McKittrick argues that this immediately undermines Rosenberg’s claim to scientific realism: “rocks are not thought to have experiences and are not considered natural individuals... by implication most of the parts of what we think of as the physical world, are less than fully actual, and have one foot outside of the natural world” (McKittrick 2006, p19). “So apparently, causal production happens in this non-natural realm of mere possibility” (*ibid*, p6);

the base from which higher-level beings such as people are formed does not exist in nature... It’s very puzzling to think that we are made up of things that don’t exist... According to his subtitle, Rosenberg’s book is “Probing the Deep Structure of the Natural World,” but the deeper he probes, the less natural it gets. (*ibid*, p5)

This is a little unfair to Rosenberg, as he is keen to link his ontology, in which temporal and spatial relations supervene on a non-spatiotemporal “causal mesh” of interconnected modal constraints to the latest theoretical physics (2004, p28; see also p213-4, n8). In fact he has good company among string theorists, whose objects of study, in addition to subtending spatiotemporal relations, seem to inhabit a no-man’s land between concrete and abstract, purely mathematical entities.<sup>47</sup> As the physicist Sten Odenwald (2002) sees it, physicists nowadays study abstract objects (p182), whose shapes “are motionless” in themselves, but “imply motion” (*ibid* 180) in those rare cases when their structures correspond to observable particles (which by and large they do not; *ibid* 177-8). McKittrick’s complaints may simply be based on an outdated conception of what it is physics studies.

---

<sup>47</sup> See Greene (2000); Montero (2001).

Nevertheless, McKittrick has noticed a tension in Rosenberg's relation to science. On the one hand, he is keen to connect his modal realism to the orthodox "many worlds" interpretation of QM (Rosenberg 2004, p288, n8), whose "layered decoherence" into the world we experience "is too complicated for *current* physicists to model" (*ibid*, p240, emphasis added). He draws heavily on neuroscientific data to argue that the brain's functional unity is a "natural individual", and seems optimistic that biology, once equipped with his ontology, could make reasonable guesses as to what natural individuals there are in the animal kingdom (*ibid*, p299-300). On the other hand, his claim that "To be actualized, an individual must take on one, and only one...potential state" (*ibid*, p158), in conjunction with his insistence that what actually constrains possibility space isn't up to science to discover (*ibid*, p218), gives Rosenberg license to ignore scientific evidence for functional or explanatory disunity, such as that surveyed in the previous section. For instance, William Wimsatt, a philosopher of science discussed in more detail in the next chapter, argues that it is just part and parcel of biology for questions of causation and functional unity to have more than one definite answer (Wimsatt 2007, p181-2). Rosenberg could reply that of course there are many individuals at lower levels of organization in the body, but only one comprising all subjective qualities, but this seems gratuitous, and he also wishes not to rule out the possibility of split subjects (Rosenberg 2004, p255). But this then introduces scepticism back into Rosenberg's account: how does he know the receptive and effective carriers instantiated in the body coincide with the deliverances of introspection? His entire argument for humans being natural individuals comes from the introspective revelation of having a unified consciousness, and panexperientialism (which he endorses; 2004, chapters 5 and 6) can accommodate illusionism about this.<sup>48</sup> The motive for identifying phenomenal

---

<sup>48</sup> Rosenberg seems to admit there is a potential for disagreement between the pre-reflectively experienced properties of a receptive carrier and the reflective judgments that come to consciousness, when he writes that "the hard problem of consciousness may be two problems superposed. It may be a general problem about

properties as candidates for identification with intrinsic properties breaks down if Rosenberg's metaphysics of qualia permits misrepresentation by phenomenal judgments.

So either Rosenberg identifies his pre-objective causal mesh with the latest physical theories, and his natural individuals with biological functions, and accepts the possibility of revisionism about the place of consciousness in nature, in line with Wilkes's view; or he should admit that his position, which he calls "liberal naturalism", is actually autonomous from science. As it stands, his "place for consciousness" merely *resembles* a naturalistic philosophy, without being unambiguously either continuous or discontinuous with scientific research.<sup>49</sup> And this puts Rosenberg in the worst of both worlds with regards to mental causation, despite it providing him with a major motive to develop a new theory of causation in the first place (Rosenberg, 2004, p108-110). The theory of causal responsibility is touted as accommodating backwards causation (Rosenberg, 2004, p149-50; Rosenberg 2017, p160) and, as mentioned, the non-fundamentality of spacetime for modern physical theory, despite asserting on non-scientific grounds that bearers of causal properties are non-physical (see McKittrick 2006, p12-3). Naturalist or not, however, Rosenberg's account of how the mind causes action is still revisionist:

On his view, there is no efficient causation between levels ([Rosenberg, 2004] 281) and our physical states do not causally interact with our conscious states ([*ibid*] 265). So, mental

---

finding a basis for qualitative fields, their place in nature, and the laws governing them. Also, it may be a specific problem about how the influence of cognition can give a qualitative field the character of *consciousness*." (2004, p104)

<sup>49</sup> For example, with regards to his ontology's relation to QM, he writes: "I am not saying experiential qualities are waves. I am saying the logic of how waves behave and combine is a useful model for the logic of how experiential qualities behave and combine: they are similar in essential behaviors" (2017, p172). But the resemblance is obscure. For comparison, Esfeld warns against reifying possibility space into a real entity with causal powers, since it is "entirely mysterious how a field on configuration space could influence the motion of particles in physical space" (Esfeld 2017, p228), whereas Rosenberg regards QM as supporting the "*ingression* from a space of possibility to actuality" (Rosenberg 2004, 211). For Esfeld, though the disposition of particles to behave in certain ways is modelled by the universal wave function, this disposition is to be regarded as primitive. It marks the limit of scientific explanation, just as the properties of Newtonian particles were, and does not indicate the presence of a more primitive force acting spookily at a distance (*'the disposition of all particles'* is topic-neutral as to what lies behind their joint disposition to move relative to one another).

states are not productive, efficient causes responsible for bodily movements. The mind does not make the body move. One might hope that a theory of mental causation would explain, for example, a causal chain between my feeling thirsty and reaching for a glass of water. Since on Rosenberg's view, the only sense in which mental states are causes of physical states is as final causes, the only sense in which my thirst could cause my reaching would be if the purpose of reaching was to feel thirsty." (McKittrick 2006, p19)<sup>50</sup>

To recap: Lockwood denied the universe is 'grainy', in the hopes of solving the bonding problem. He agreed with Russellians that the universe in itself is qualitative. But the conceivability of a zombie quantum universe without awareness of qualities meant that for Lockwood experience is still epiphenomenal. He also conceded he has no idea how to account for intentionality, therefore failing to account for phenomenal judgements or awareness 'of' qualities. Chalmers conceded that fundamental physical entities need not be microscopic, they just seem that way to philosophers prioritising physics; but he went on to insist that the importance of logical supervenience to metaphysics forces us to prioritise physics anyway. Powers-metaphysics appeals to modern physical theory rather than philosophers' a priori intuitions about supervenience, challenging Chalmers' Humean ontology, and postulates fundamental entities at potentially very large scales. Building on this, Rosenberg then fills out the details (the property of nomic necessitation has a phenomenal realizer: his "carrier hypothesis"). He diagnoses the Scylla of the bonding problem, which leads to eliminativism about o-properties, and the Charybdis of the decomposition problem, which leads to unrestricted panpsychism (Rosenberg 2004, p84-90). Chalmers has criticised his solution as dualist (2017, p193-4), but a better argument –

---

<sup>50</sup> As the next chapter discusses, it is ironic that the place Rosenberg reserves for mental causation is precisely via a form of explanation that modern science did away with: the appeal to a final cause.

McKittrick's - is that it looks like idealism: Rosenberg avoids his rivals' problems by giving lesser reality to basic physical entities.<sup>51</sup>

Like Wimsatt, Ladyman *et al* (2007, ch.4) argue that there is no one answer to how something is individuated; this is just part of taking complexity seriously, not a failure of science to appreciate causal significance. For Rosenberg to deny this development in scientific explanation, and to take conscious personhood as the benchmark of a natural individual, is to make man the measure of all things. In doing so Rosenberg introduces a whole new ontology with new terminology, one that is highly speculative (as McKittrick complains, p20), and one liable to draw mockery from naturalist philosophers such as Ladyman *et al*, for whom this will be one more example of 'domesticating' metaphysics ending up just as counter-intuitive as mainstream physics. As far as naturalists are concerned, what we need is a non-reductionism that wholeheartedly takes its cue from current science: no modifications of nature are within the philosopher's purview, nor should they seek to reinterpret scientific results too much.<sup>52</sup> And for this we need to go beyond tinkering with the ontology that set up the bonding problem(s) in the first place, and consider the deeper epistemological and metaphysical issues from which the hard problem itself arises. The next chapter will take a historicist approach to the issue, setting up a physicalist alternative in Chapter 4. First I want to consider just how alien naturalism is by comparison to old-fashioned metaphysics, once again with reference to Aristotle.

---

<sup>51</sup> Rosenberg has since been forced to clarify that this isn't what he had in mind (Rosenberg 2015, p241), but, as the next chapter argues, it still isn't clear *what* he is saying, and more to the point, that insofar as Rosenberg is a scientific realist he has no basis for disagreement with Wilkes about the relevance of consciousness (as opposed to Whitehead's watered-down 'experience') to 'serious' ontology.

<sup>52</sup> Whether this is a realistic constraint for the philosophy of consciousness is another issue, of course. But the important point is that Rosenberg himself identifies as a naturalist and a scientific realist, and so is vulnerable to the objection that he fails to walk the walk.





## Chapter 3:

### Historicism and Naturalism

#### [3.1] Aristotle on Matter

Wilkes and Heil both want to render Aristotle's ontology commensurate with physicalism (or at least scientific realism, in the case of Heil's quasi-panpsychist metaphysics). Correlatively, they downplay the extent to which Aristotle ultimately diverged from it. I say ultimately because it is only in the final analysis that Aristotle's worldview begins to look strange; when aspects of his explanatory strategy are taken to their logical conclusions. These conclusions concern the nature of matter, the mind's relationship with the body, and the existence of God – conclusions which are inextricable from one another. While the attack on Humean supervenience could plausibly support the irreducibility of organic life to its parts, and appeal to Aristotle in doing so, it cannot follow Aristotle's subsequent likening of the universe to a form of life or mind. Of the philosophers discussed so far, only Rosenberg comes close to reframing Aristotle's complete metaphysics with reference to modern science, but as I argued, his account has problems at precisely those moments when the question of scientific realism comes into view.

The first obvious respect in which Aristotle diverges is in his treatment of matter. Section [1.5.1] briefly touched on Aristotle's account of change and persistence through

appeal to matter and form, and noted that there is, strictly speaking, no such thing as matter in itself on this picture, only enformed matter. Matter may lose its substantial form, but it will always acquire a new one straight away. As Shields (2013) puts it, “[t]o be is to be some F; nothing simply is” (p193). This is especially clear in Aristotle’s treatment of organisms, where the temporally extended processes associated with living make essential reference to organic matter, and vice-versa. An animal’s substantial form just *is* the animal (Shields 2013, p183-4), a metaphysical simple in which predicates such as growth, nourishment and so on are grounded.<sup>1</sup> Hence an animal’s parts are fundamentally alive, as predicates of the whole, and so persist only homonymously after death; a dead eye is not really an eye any more.

It might seem obvious to us that it is matter which persists through time and which gains and loses form. As Shields (2013, p276-9) argues, Aristotle may well have changed his mind about this. In his *Categories* Aristotle identified Socrates as a compound of matter and substantial form; later, in the *Physics* he wondered (without answering his own question) whether substance is matter or form (*Physics*, 191a19–20). Later still, Aristotle concluded in *De Anima* that substance is form, and hence that what persists through change is not even an approximation of the kind of material stuff we believe in today. One might have assumed that denying reductionism could give physicalists leeway to go along with this; however, given Aristotle’s explanatory schema, and the central role substance plays therein, it soon becomes clear that identifying substance with form takes us far away from reality as we conceive it today.

Section [1.5.1] noted that Aristotle built his ontology in response to worries about the existence of time generated by sceptics such as Zeno. In response to the latter’s insistence that time’s infinite divisibility renders motion impossible, Aristotle argued that

---

<sup>1</sup> Shields likens substances to linguistic phonemes, which, while not strictly indivisible, are “basic units relative to a context of appraisal” (2013, fn.14 p228).

time is only potentially infinite in extent and divisibility, because time is a quantity of change. Criteria for what changes can be measured are needed before the amount of time taken has an answer. The persistence of objects – substantial forms – supplies that criterion: “the kinds of motion and change are as many as the kinds of being” (Aristotle, *Physics*, 200b32–201a9). This relates back to the claim that not just anything is possible; *how* matter is actualised delimits how it can change. It rules out matter ‘as such’ being subject to change; only substances change, in ways appropriate to their dispositional essences, as Heil would say. There is no change in-itself, because matter itself only provides the *potential* for change. Matter accounts for the possibility of losing a form, but not, by itself, the persistence of form; this is accounted for by the substance that a thing persists as.

Nevertheless, important traces of Aristotle’s predecessors’ insistence that reality be unchanging survive in his account of change as “the process of actualization, and not the product or state resulting from that process. By the time we reach that stage of actualization, the change is complete” (Shields 2013, p240). On the one hand, then, the most actual, i.e. reality as it truly is, is unchanging, and since only the actual can effect change, the ultimate source of change must be the so-called “unmoved mover” – identified by later commentators with the Abrahamic God. On the other hand, this causal influence cannot be understood as just setting things in motion,<sup>2</sup> because becoming fully actualised, and thus unable to change further, is also the end towards which things tend.<sup>3</sup> Causation is not therefore simply a temporal association between events, but an ontological relation between things. A thing “owes its continuing existence “here and now” to a cause above it in the hierarchy of being” (Moevs 2005, p28); this was also Rosenberg’s approach to causation,

---

<sup>2</sup> As in material causation, “That from which an entity comes to be” (Shields 2013, p52), or in efficient causation, wherein a form is imposed.

<sup>3</sup> As in the formal cause making a thing be what it is, and, especially, the final cause for the sake of which it exists.

as we saw in the last chapter.<sup>4</sup> Aristotle's God is the ultimate *reason* for why anything is in motion (exists) at all; it provides the grounds of actuality from whose standpoint all causation could be viewed as a single 'act' of the unmoved mover. Christian Moevs – discussed further below – tracks this “identification of God with pure existence” through antiquity in the hands of Islamic, Jewish and Christian commentators on Aristotle, culminating in Thomas Aquinas's slogan that existence is an act, not a thing (Moevs 2005, p30). In a further anticipation of Christian doctrine, Aristotle claims “the unmoved mover causes motion insofar as it is an object of love” (*Metaphysics*, 1072b3–4), with love glossed by Moevs (2005) as “the experience or recognition of the other as oneself” (p69).

It may have initially seemed as if matter must be explanatorily and temporally prior to form because it gains and sheds forms. However, for Aristotle, form is “actuality, and matter potentiality; and it is important to stress in this connection that actuality is prior to potentiality” (Shields 2013, p301). So as things are actualised, or progress towards their final state in which further change is impossible, they lose their matter (since being material harbours the possibility of change). In a sense of course this is a matter of realising (in both senses: actualising and recognising) what was already the case, since properly speaking only pure form is actual. Matter is denigrated twice over in Aristotle's system. Firstly, by the role assigned to proximate or relative matter, that which underlies a particular existing thing. As Moevs puts it, “certain materials are more suitable for a particular use than others” (2005, p45); a wax can be “disobedient” to the form impressed on it. Since things can only be known insofar as they are formed, however, and since substances are the grounds of their parts and hence the four elements, it is not the case that proximate matter persists independently as those elements. Rather, form and proximate matter are “somehow” one (*Metaphysics*, 1045b21); bearing in mind that form persists while its matter changes. The

---

<sup>4</sup> A commitment that is absent from Goff's variant of layered monism; hence why I described it as dualism.

matter of such a compound amounts to the privation of the substance which a thing is: “not only is (proximate) matter responsible for all imperfection in nature, it is also the principle of non-intelligence or non-intelligibility in reality” (Moevs 2005, p45). Proximate matter is a roadblock to actuality.

Secondly and more importantly are arguments for the incoherence of ultimate, unqualified or unformed matter, which is treated, like time, as only potentially infinitely divisible, and not objectively either divided into fixed units or an actual infinitesimally fine ‘gunk’. But whereas forms helped shore up the reality of time against scepticism, here their necessity only underscores the unreality of matter construed independently of form, as such a construal proves to be impossible. A substance “is both some determinate thing and something actual: but matter thus construed is in itself nothing particular” (Shields 2013, p304). Nothing persists except as this or that, but formless matter could not even be quantified, let alone measured, being without positive characteristics, “simply, indeterminate” (*ibid*, p305).

A final and crucial contribution of Aristotle is his denial that the potential for change (which matter supplies) could possess independent existence. This marks a break with Plato, as Moevs explains:

Plato’s *Timaeus* (49–52) speaks of an ultimate, undefinable, virtually unknowable receptacle of all things, “an invisible and shapeless kind of thing,” an eternal space “that provides a seat for all things which come to be” ... Quite naturally, subsequent thinkers, beginning with the Stoics, identified this receptacle with Aristotle’s prime matter. This Aristotelian-Platonic conflation was then identified in turn with the earth created as a formless void. (2005, p40)

The Platonic reification of prime matter enjoyed a brief resurgence in the Middle Ages; “some Scholastic philosophers, especially in the Augustinian-Franciscan tradition, thought that “first matter” could exist separately in time, either through having some slight actuality or determination (e.g., Bonaventure) or, since God could do all things, as pure potentiality (e.g., Duns Scotus)” (*ibid*, p41). This led to the thought that only what is abstractly possible is truly real, leading in turn to the privileging of logical possibility in metaphysics and, ultimately, to Lewis’s modal realism.

For Aristotle ultimate reality cannot be construed in these terms. Yet a version of it seems to persist in his treatment of mental representations. Our capacity to experience forms is argued to entail an infinite plasticity of mind, whose ability to take on and in some sense actualise and become identical with the essence of anything rules out its being identified with the limited powers of any proximate matter such as the brain. Prior to thinking, then, mind seems very much like prime matter, a pure receptacle of infinite potential; and as Shields (2013) notes, it hardly seems more sensible to claim to the contrary that “the mind simply does not exist before it thinks, that it somehow suddenly pops into existence just at the moment thinking gets underway” (p352). Moreover, when it is thinking, this “active mind” is said by Aristotle to somehow share in the activity of being that is the unmoved mover, in controversial and cryptic remarks made in section 3.5 of *De Anima*. And doesn’t purely potential intellect share its being with all things qua matter, and therefore without having to start thinking or becoming them?

Catching out these claims fell to Aristotle’s more overtly religious commentators of subsequent centuries, whose conclusions, Moevs argues, were ultimately to be mythologised in Dante’s *Divine Comedy*. I turn now to Moevs’ account of that tradition.

### [3.2] Aristotle on Mind

Christian Moevs' 2005 study situates Dante's poem at the peak of medieval Aristotelianism, the conclusion of philosophical debates engendered by Neoplatonists, Christian mystics such as Meister Eckhart and Pseudo-Dionysus, as well as prior generations of commentators on Aristotle; most especially Averröes in the Middle East and Thomas Aquinas in the Latin West (and Dante himself, in his philosophical writings prior to the *Comedy*). This is somewhat unorthodox, for while the latter philosophers could be regarded as participants in mainstream theological debates, with Aquinas in particular earning a reputation as the great synthesiser of Aristotelian and church doctrine, mystics such as Eckhart were often viewed with suspicion, and it is the mystical aspects of theology that interest Moevs, since these pertain to the relation between consciousness and fundamental reality. Indeed, Dante's *Comedy* almost suffered condemnation for his claim that the poem revealed the true meaning of existence and of Christianity, and even read the mind of God (Moevs p175); the latter being construed along the lines of Averröes's monopsychism, wherein all minds and God's mind are one, a claim famously attacked by Aquinas (Stone 2000). Since the metaphysics of Dante's *Comedy* will be taken in this chapter to exemplify a worldview which was not displaced until the scientific Enlightenment, the poem's overall relevance to the medieval *zeitgeist*, when not construed as merely a work of fiction, needs defending.

There are two lines of response: one, which is taken by Moevs, is to emphasise the continuities between all the thinkers just mentioned and downplay their differences. The other is to concede that Moevs' syncretism can at best only capture an aspect of medieval thought, but to insist that it is an important one. The thrust of Moevs' polemic is that those who would deny Dante's message, whereby we can come to know the nature of God in our lifetimes, risk reducing religion to dogma, and ignore the purely philosophical motives for

introducing the unmoved mover in the first place. More importantly for current purposes, is that in claiming that grasping the true nature of experience solves the riddles Aristotle left outstanding in *De Anima* 3.5, Dante presents a solution to the mind-body problem which could only be endorsed during the period when Aristotelian ontology was in vogue. That other philosophers of greater stature, such as Aquinas, made more modest use of Aristotle's ideas and insisted that full metaphysical understanding was beyond our grasp,<sup>5</sup> does not detract from the interest that such an all-encompassing framework has. Unlike dualism, which was the received wisdom in Western philosophy prior to the rediscovery of Aristotle in the 12<sup>th</sup> Century and would make a comeback thanks to Descartes,<sup>6</sup> Dante's position does not obviously conflict with the causal closure of physics. Indeed, Moevs' study of the *Comedy* ends with a review of overlaps between Dante's metaphysics and contemporary physics, which he interprets to have done away with time, space, and matter (p187-192). Rosenberg's ontology has strong similarities to all this (as I argue below), and bears out Moevs' contention that the medieval explanatory framework once again deserves taking seriously.

Perhaps the easiest way to appreciate Moevs' defence of Dante's mystical Aristotelianism is to compare it with Kant's transcendental idealism, but, as befits a pre-modern outlook, with Kant's constraints on knowledge lifted: intellectual intuition is supposed to be within our reach. Moevs argues that the "notion of matter is grounded in the duality between the subject (immaterial form or intelligence) and object (spatiotemporal

---

<sup>5</sup> "Aquinas thought, at least for a time, that in this life one could not experience Intellect-Being in itself" (Moevs, 172-3). Brian Davies' (1992) study of Aquinas emphasises that for Aquinas we don't need divine guidance to know anything, don't always already know God exists, and must observe world to know God – there is no mystical travel inwards, only an intellectual inference of a first cause in Aquinas's proof of God (about which he concedes the possibility of reasonable disagreement). Moevs (with Dante) seeks to emphasise a sense in which belief in God, being inextricable from belief in the reality of conscious experience, is undeniable.

<sup>6</sup> Early dualism was informed by Neoplatonism and Saint Augustine; Stone (2000), p46. Stone emphasises (*ibid*, p34-5) that Augustinian dualism was already more Aristotelian than its Cartesian successor, but given the subsequent controversy over Augustine's relation to Aristotle in Stone's narrative, I shall set that aside.



form) of experience” (46). Since material objects are only fully actualised and knowable as form, this makes their materiality an upshot of a failure, on the part of the subject, to contemplate them sufficiently. Intellect is blocked from full contemplation of the forms by the “disobedient” proximate matter upon which it is stamped (the organism that one is). It seems to follow that we are stuck with apprehension of existence through concepts, rather than knowing things in themselves directly; a predicament that makes our form of knowledge irremediably “discursive” or non-metaphysical for Kant (Allison, 2004, p.xiv). The alternative, intuitive, grasp of objects, Allison notes (p.xvi), would involve unmediated and non-conceptual access to objects, and thus no subject-object duality at all; “God knows creation as His own ideas, which are identical with Himself” (Moevs p46). The latter is thus necessarily self-contemplative, i.e., perceiving all experiences as variations of self-perception.

The possibility of our possessing this extraordinary capacity ourselves is virtually a necessity in order for Aristotle’s framework to make sense; without a latent potential for *nous* to overcome its material limitations, its relationship with the external world looks precarious, and sceptical positions, taken in different ways by Descartes and Kant, begin to recommend themselves. Indeed, according to King (1994), the failure of medieval philosophy to explain how such a thing is possible was precisely what led to the abandonment of Aristotelian psychology in favour of substance dualism. King focuses on the problem of relating sense-perception, which is inextricable from proximate matter and hence imperfect, to *nous*’s grasp of the eternal forms, which it must somehow “transduce” from the available evidence given to sensation. Common sense unites the incomplete ‘sides’ or aspects of a sensed thing into one thing, but a further faculty is needed in order to render the latter intelligible as an instance of a universal, i.e., to become aware of sensations as representing objects. Since objects are the agent cause of the sense faculties, a distinct agent cause is

necessary to actualise the potency of the understanding (extracting the universal from the particular) as opposed to sense. This is the 'maker mind', or light, provided by *nous*, which King interprets rather literally as a mental module for transforming cognitively impenetrable sensations into symbols. Aquinas and Duns Scotus thought transduction happened via abstraction; so-called illuminative accounts (Bonaventure, Matthew of Acquasparta, and Henry of Ghent) differ in that knowledge of the forms is supposed to be implanted by God. The former account has the problem of suggesting that forms are already implicit in sensory states, so the introduction of a separate agent faculty is gratuitous, while the latter implies that sense-perception is a hindrance to true knowledge. In neither case are the relations between universals and particulars, ignorance (or privation) and knowledge, non-conceptual and conceptual mentation, perception and justification, etc, actually accounted for; all variations of a mind-body problem which Aristotle never satisfactorily solved. It is no surprise, then, that Cartesian philosophy responded by making the mind-matter distinction fundamental and inexplicable, collapsed the faculties of (conscious) sense and understanding into one, and borrowed from illuminative accounts in defending innate ideas (King 1994, p124).

The mystical solution to this dilemma attacks it at its root, in the very positing of a distinction between subject and object which sets it in motion. Even if this solution smelled of heresy whenever it recurred during the middle ages (which explains why it was never going to win a popularity contest with the dualist solution), it has the metaphilosophical advantage of giving full weight to the Aristotelian attack on prime matter, a point which King's functionalist account of the problem downplays, since the point of identifying reality with form was to *avoid* postulating a non-cognitive ("cognitively impenetrable") level of reality which could not be contemplated. The ultimate transducer is one wherein the medium or veil between us and universals is dissolved in immediate experience (Moevs

2005, p82). Denying such a thing leads to dualism (with the problem of causal interaction) or Kant's phenomena/noumena split (with the problem of how noumena can have causal force), which was likewise predicated on an absolute distinction between sense and intellect (Yovel 2018, p8-9). And while "some Kant followers" suggested the latter "branch off from a secret and unknowable common root", that "source can only be intellectual intuition, and whoever affirms the existence of such a source is already claiming to know and to use it" (*ibid*, p10). This claim was bound to sound impious in the middle ages, and strikes us as arrogant even today, so its plausibility hinges on being less radical than it sounds.

According to Dante's mystical solution, active intellect just *is* the passive (sensory) intellect considered in itself, in its grounds of actuality; there is no causal relation between a real thing and its reality. That all things are therefore identified with acts of thought raises the puzzle of how the potential intellect could play an explanatory role in accounting for ignorance, and how it could even be distinguished from any other form of (supposedly unreal) potential. "Why is not every object *nous*, intelligence, if it can exist intermittently as mind? Or, from the other side: how can form be only intermittently an intelligence? What makes it so, if it is not always so?" (Moevs 2005, p52). Proximate matter cannot supply the answer if it is in reality always already an act of the active intellect. But thought always starts somewhere – from a particular, actualised form. Mind does not exist prior to thinking because nothing exists prior to the self-contemplation of the unmoved mover, of which each of point of view is supposed to be a variation. Individual perspectives couldn't even start anywhere without possessing the further advantage (not granted to just any finite thing) of being able to assume any form. Mind as such thus exists in an unqualified fashion. A subject's finitude, prior to having learned about the world, does not materialise it – our starting position of ignorance is still a limited form of divinity because finite beings only

subsist relative to mind, but not the other way around (to have sensory experience is still to think).

The perishability of memory is crucial here (*ibid* p75). It seems to play a double role as both the instigator of thought (via impressions retained from the senses) and a veil preventing full knowledge of things – universals – in themselves. But the latter never needed to be traced in memory because we ‘always already’ know them; we don’t need to remember anything *in particular* to embody our true nature in and as form. Being unqualified, our true nature is therefore strictly speaking neither spatial nor temporal (*ibid* 136). But this is not to give any independent reality to non-spatiotemporal stuff; denying such a thing is a crucial point of contention between Aristotle and believers in prime matter or Rosenberg’s possibility space. On the contrary, reality cannot be other than or prior to its appearance(s). For Aristotle the actual is always qualified as the presence of some particular changing thing(s), but conversely, the process of actualisation is a qualification of the actual (the now), which is itself unchanging, and which gives a window into the single, infinitely extended moment in time of the act of creation. The present circumscribes the extent to which things are actual, and thus capable of change. Now is always some now, some particular moment in time; and here likewise some particular place. But their perpetual qualification, the always-already contingent and transitory content of time and space, is a timeless property thereof, their form which is itself unqualified. Finitude or privation in-itself *is* the infinite, and the self-same universal or divine subject cannot but be given to itself as an infinity of differentiated points of view; as self-contemplation, the unmoved mover cannot but be self-identical by virtue of its self-differentiation. Moevs emphasises the appropriateness of Christian symbolism, specifically the crucifixion and the solar eclipse which supposedly accompanied it, whereby the very moment at which the divine is most obscured paradoxically reveals it for what it is. But he also notes a continuity with

Wittgenstein's influential image of the visual field as both finite and limitless,<sup>7</sup> which he reads as not just describing the way in which we represent the world, but an essential property of the world itself.

The mystical solution to Aristotle's mind-body problem ties up the loose ends left in *De Anima* with regards to the relation between the active intellect, or *nous*, the ensouled body of hylomorphism, and the unmoved mover. Its solution is radical, identifying mind, substance, and fundamental reality in terms of one another. However, in keeping with Aristotle's denial of a separate world of forms, this solution must also ultimately conclude that the world of appearances is not other than the unmoved mover, but a way in which it must necessarily appear – to itself, no less – and thus may be more plausible than it first sounds. As the culmination of Aristotle's metaphysics, this claim would enjoy popularity in the Middle Ages largely thanks to Averröes and his European exegete Siger of Brabant (Stone 2000, p42), but was attacked by Aquinas and later condemned for heresy in 1270 (*ibid* p52). While Dante's *Comedy* sought to reconcile Siger and Aquinas (*ibid* p55), and remains widely read, its reputation is as a work of fiction, which is how Dante himself escaped condemnation at the time (Moevs 2005, p175). The significance of the *Comedy's* metaphysics to our own times therefore depend on its applicability to our modern worldview, which is still shaped by Descartes' rejection of Aristotelian philosophy. As Sarah Patterson's re-evaluation of Descartes<sup>8</sup> shows, a crucial part of his legacy, which has nothing to do with atomist or 'Humean' ontology, is still with us. The rest of this chapter will therefore question whether a return to Aristotle, of the kind Moevs recommends, is even possible.

---

<sup>7</sup> Moevs 2005, p170; Wittgenstein (1974), *Tractatus*, §5.631-5.633; §5.641.

<sup>8</sup> "How Cartesian Was Descartes?" (2000).

### [3.3] Descartes versus Aristotle

Patterson disputes the ordinary characterisation of Descartes as a sceptic and dualist whose neurotic need for certainty in the face of radical doubt leads him to discover that the existence of his mind is not subject to the same doubts as the material world, and so cannot be a material thing. This characterisation, which Patterson calls the “Cartesian” interpretation, sees Descartes as leaving us with problems that would have baffled the ancient Greeks and which likely cannot be solved: the epistemological problem of proving we have knowledge of the external world in spite of knowing with certainty only the contents of our minds, and the metaphysical problem of showing how mind and body interact. The “Cartesian” Descartes has a patchy reputation. Since the problems he left us with are insurmountable, and in any case contrived by Descartes’ overly strenuous standards of justification, modern philosophers conclude that they can and must be set aside. Chalmers and Goff both certainly refuse the epistemological challenge, because they deny that a subject should necessarily know *why* their knowledge is justified, or whether they have knowledge at all.

Patterson writes that “[t]he Cartesian interpretation leaves Descartes with a conception of mind as the repository of perceptual appearances and of the external world as the repository of the familiar objects we used to believe those appearances revealed to us” (Patterson 2000, p76). But Patterson argues that Descartes does not ground his knowledge of what there is in the world upon the incorrigibility of what *appears*, and denies that the external world *resembles* those appearances. So on this interpretation Cartesian scepticism should not lead to Berkeleyan idealism (which discards the external world as explanatorily irrelevant); since he doubted the deliverances of perception, Descartes could not be seeking to reconstruct knowledge of the external world by appeal to sense-data. Instead, “he uses

sceptical doubts strategically to counteract common sense and Aristotelian preoccupation with sensory perception and prepare the way for a purely *intellectual* understanding of body, mind and God, based on innate ideas....Descartes' chief aim is to provide metaphysical and epistemological foundations for his *mathematical physics*" (*ibid*, p77, emphasis added). The success of rationalism in science gives *inductive* reason for doubting the deliverances of the senses. That the ideas we have of mental, physical and divine attributes were thought to be beyond doubt only provided a further motive for breaking with Aristotle's empiricism. "[S]ceptical doubts are used chiefly as a strategic device" for weighing the merits of sensory versus intellectual understanding. The aim is to explain what it is our knowledge is of, rather than the achievement of perfect certainty (*ibid*, p106): the Cartesian universe is apt to be understood primarily by rational rather than empirical sources of information. Descartes' remarks that mathematical truths are unaffected by doubt are undermined, Patterson argues (*ibid*, p87-9), only by the possibility of our having been made imperfectly, a possibility of which Descartes is far less sure than the possibility that he is dreaming.

For Aristotle and his medieval followers, nothing could be more obvious than that knowledge must start from experience, or sensation, by which the forms or universals in nature enter (are imprinted on) the mind. Form and matter are always conjoined, so to apprehend a form at all the mind must turn to its instantiation in an empirical individual (*ibid*, p79-80). Descartes argues that this conception leads to the identification of mind-independent nature with what can be pictured or imagined, a habit of 'common sense' which can only be remedied by entertaining radical doubts about the senses (*ibid*, p90). Having dislodged our old habits, we will finally be in a position to distinguish between the objects of empirical knowledge and the far more valuable objects of purely intellectual understanding. For instance, the persistence of wax through its modifications is not known by reading off an unchanging universal common to any possible changes it could undergo,

since it has more possible forms “than the imagination can encompass” (*ibid*, p94).<sup>9</sup> And for Descartes, everything persists in the same way: in getting ‘behind’ appearances to the numerical identity of the wax (ultimately its chemical composition) the mind is not discovering a unique essence, distinct from all others, but merely a modification of extended substance as such, the basic matter which is the object of all scientific enquiry. “The variety in corporeal things, including the variety in living things, is due not to the different natures by which different parcels of matter are informed, but to the shapes and motions of the parts of matter itself” (*ibid*, p79-80). Descartes therefore challenges Aristotle’s argument against the conceivability of prime matter, a position bolstered by the return of atomism in 17<sup>th</sup> Century physics (which contrasts to the potentially infinite divisibility of matter for Aristotle). The plurality of natural forms we perceive are to be replaced, Descartes argues, with just three forms known innately to the intellect as completely distinct from each other: matter, mind, and God.

Patterson interprets the distinction between the latter three ideas as evidence that “Descartes does not conceive of the external world (what is outside the mind) as a realm that is metaphysically or epistemologically unified” (*ibid*, p103), given that we know about the three substances “in different ways”. By contrast, the Aristotelian world was understood a posteriori through seeking greater unity, arriving at a proper understanding, and unification, of Descartes’ three ideas only at the end of a long, laborious process of ‘recollection’ (Moevs 2005, p52-3), at which point the true nature of things will be *instantiated* by the subject themselves. Greek and medieval epistemology were directed outwards: for them the “human intellect knows first the nature of a material thing, and only second its own act by which the object is known. As Aquinas puts it, the intellect understands a stone (the first act) and understands that it understands a stone (the second

---

<sup>9</sup> This is somewhat compatible with the persistence of the wax still being apprehension a form, but is more in line with Platonic than Aristotelian metaphysics. See below.



act, parasitic on the first)" (Patterson 2000, p92). Even the mystical turn inwards, and with it the discovery of the "thought that thinks itself" was construed above all as a consequence of performing good deeds, whereby we identify ourselves with others (Moevs 2005, p68-9). For Descartes the recursive intellect is the *first* thing we can know, the easiest to grasp; clarity and distinctness are properties of individuals, not totalities, the wider context things exist in can be bracketed, and the *cogito* is solitary rather than sharing the form of the divine. Aristotle's holism is not like modern science, which often proceeds piecemeal rather than holistically; for instance, Bechtel (2007, p10-11) argues that biologists posit narrowly context-specific mechanisms, rather than laws, to explain organic processes.<sup>10</sup>

All of 17<sup>th</sup> Century science seemed to weigh against the old picture of form and matter being intermingled: the possibility of isolating objects of experimental significance, the success of Newtonian physics, and the irrelevance of the secondary qualities, strongly suggested that matter has an existence separate from that of form. And "to Aristotelian eyes, the claim that matter is subsistent is far more striking than the claim that the intellectual soul is subsistent" (Patterson 2000, p96). Matter had no independent existence for Aristotle. By modern standards, and despite Moevs' aforementioned claims to the contrary, this would make Aristotle an idealist. Moevs takes idealism to mean a response to scepticism – an epistemology, not an ontology. And since scepticism is supposed to be bogus, idealism must be as well. But he should surely be aware how similar the subsumption of matter by form is to G.W.F. Hegel's *absolute* idealism, whereby matter is ultimately 'sublated' into mind: "reality may be described in two interchangeable ways: as being, or as

---

<sup>10</sup> Which is not to deny that *seeking* unification motivates good science; this is just as true now as it was for Aristotle, who after all made the first efforts at scientific explanation. However, situating a thing in the widest possible context is no longer prerequisite for fully understanding it; on the contrary, objects of enquiry must normally be isolated from the environment in order to be studied in detail. Moreover, since understanding is now arrived at from the bottom up, as it were, there is no longer a guarantee that God will be waiting for us at the end of enquiry, whereas for Scholastic Aristotelians the highest being was supposed to be known in advance as the source of the 'great chain of being' into which all other phenomena could be fitted.

intelligence-intelligibility” (Moevs 2005, p57); moreover, “in an absolute sense, only God is; all else shares” (*ibid* p58). If this were truly a third way *between* the binary oppositions mind and matter then matter would have to be granted much more reality than simply that of privation. Even the idea that the universe has a preferred way of counting objects (forms in the great chain of being) seems, from our perspective and Descartes’, like projecting human criteria of sense-making onto the cosmos; an idealist move. This compromises efforts to reconcile the old Aristotelian worldview with modern science, as I shall now argue.

### **[3.4] Science Versus Common Sense**

#### **[3.4.1] Rosenberg’s Ambiguities**

Rosenberg thinks it’s just a coincidence that the upshot of his thoughts about causation are strange properties that resemble those of consciousness (Rosenberg 2004, p270). But the intuitions supporting his metaphysics are similar to those informing Aristotle, and the upshot is similar too. His insistence that every cause be predicated to an individual is just Aristotle’s subject-predicate metaphysics, as is the thought that we cannot predicate properties to objects without talking about their causal powers (which he thinks plugs consciousness directly into causation). His concept of causality as a relation of ontological dependence between events, rather than a temporal relation between things, would have been familiar to medieval philosophers (and to the Neoplatonists who influenced them). His layered ontology of increasingly higher-order individuals actualising their own subvenient parts could just as well have been worded in Neoplatonic jargon as a “descending series of emanations or hypostases” (Moevs 2005, p60). Rosenberg somewhat obscurely suggests, near the start of his discussion of causation, that “for all we know, existence might be

something toward which all things tend” (Rosenberg 2004, p151). This was an orthodox view among medieval Aristotelians, who believed God realises all possibilities, and that all things ‘strive’ toward actualisation (Moevs 2005, p43; 54-5).

Most importantly, for both Rosenberg and Aristotle matter is only determinate, only has existence, insofar as it is situated in the context of a form; for Rosenberg this means that the mind-independent determinacy of *parts* of our bodies, and of the wider environment, is so only “from the perspective of a higher-level individual it helps to constitute” (Rosenberg 2015 p241). For this reason, even the mystical aspects of Aristotelian thought overlap substantially with Rosenberg’s own appeal to meditative states of “contentless openness”, which he regards as exposing the experiencing subject as a “receptive carrier” of causal constraint (see previous chapter). This, as Moevs argues, amounts to assuming a God’s eye view on things: to know God is not a relation between oneself and a peculiar objective entity, but a kind of *experience* in which objectivity dissolves, in the revelation that the ultimate form, which contains all others, is at the same time a subject or thought; a “thought that thinks itself” (*ibid*, p17). Rosenberg proposes something much along these lines as follows, albeit with shades of Eastern rather than Western symbolism:

At the bottom—level zero—there is a Turausky Fundamental Tone, the white noise of the universe, in which “twenty thousand experiential tones” are superposed into a single experiential simple. This contains within it the potentials for all possible carriers of effective constraint (physical properties when viewed from the outside). Within the framework of [the Theory of Natural Individuals], we view Turausky’s Fundamental Tone as the carrier of a universal level-zero effective property, containing within it the potential for all the level-one effective properties in the world. It is the canvas of causation on which the picture of the colorful world is painted (see [Rosenberg 2004], 143), the abstract metaphysical background against which the jewel of the actual world is set. (Rosenberg 2017, p174)

Turausky's claim, further developed by Rosenberg, that the sum of all experiences is nonexperience, can be situated squarely with mystical traditions in which paradoxical claims of this kind are commonplace.<sup>11</sup> Moevs traces variations of it from Plato through to medieval theology: that the one and the many, everything and nothing, are somehow the same (Moevs 2005, p69-70).

The upshot of all this, as Moevs sees it, is that in the tradition of form-matter metaphysics, "ontologically the brain depends upon consciousness, and not vice versa" (*ibid*, p37). This certainly seems to be Rosenberg's view, since it makes organic life a paradigm of natural objecthood, and conscious selfhood a paradigm of that. But the same idea was arguably present in Russellian monism all along: Goff argues that to imagine a brain in itself would be to imagine a subject (Goff 2010, p128-9). Matter as we normally understand it – as inert, mindless stuff – only exists for panpsychists relative to our limited intellects; in reality, all there is is mind. And it is easy to see how this view could eventually turn theological: if to be a unity of experience is to be a subject, and we believe in the unity of nature, then nature must be a subject too – precisely where Goff has ended up with his turn to cosmopsychism.

The Cartesian project rules out the kind of deep correspondence between reason and nature sought by Aristotelian and medieval mystics. Indeed, the impossibility of such a correspondence seems to be exactly what guarantees nature as a mind-independent order in the first place, in contrast to the idealistic implications of identifying nature with form, because for Descartes mind-independent objects do not literally enter into *nous*, but simply act causally upon it. Strangely, Rosenberg's metaphysics continues to pay lip service to the

---

<sup>11</sup> "Because the depth-mystical experience is free of differentiated features, the state of "pure consciousness" is sometimes characterized as a state of *unconsciousness*—i.e., the meditator is in some sense awake but not conscious (Pyysiäinen, 2006)" (Jones, 2016, p22). Jones is sceptical of this interpretation, wondering why "contemplatives [should] undergo long years of training to achieve a state that could readily be achieved through a swift blow to the head with a heavy blunt object" (*ibid*). I seek to assuage this worry in ch.6.

same distinction between mind and matter, transposing it into the separation between receptive and effective properties, and hence between the subject, or perspective, and the contents of consciousness, or qualia. Recall that in [2.2.1] Rosenberg was shown to be a panexperientialist, for whom qualia are observer-independent entities. He distinguishes (2017, p174) between Turausky's Fundamental Tone and "experiential quanta" which could be thought of as proto-subjects. For him reality is first of all a pre-reflective feeling rather than a self-conscious knowing; forms (subjects) are derivative of prime matter (the modal background or Turausky's Tone), which has some independent existence of its own. Of course there could be a sense in which the nothingness of a subject (Rosenberg's receptive carrier) without qualia coincides, at the root of reality, with the nothingness of Tuarusky's self-cancelling plenitude of qualia (the foundation of effective carriers); perhaps ultimately container and contained are one, just as they were for Moevs' mystics. But the truth is that Rosenberg never explicitly argues for a coincidence between form and matter, and should not, albeit for two more or less incompatible reasons. Firstly, preventing the reduction of qualia to functional (i.e., *formal*) properties, instead identifying them with "cognitively impenetrable" sense data (King 1994, p110-1) whose relationship to, indeed knowability by, the intellect is fraught with difficulty (*ibid*), as we saw in [2.2.1]. Secondly, however, the very independence of qualia to mind – their role as quiddities or essences of external objects – prevents the reduction of reality to ideas.

As already discussed, Rosenberg grounds his metaphysics in an extraordinarily vague postulate, "abstract modal realism", a contextless background of possibilities out of which all things are "actualised", through a process of contextualisation Rosenberg calls "ingression" (2004, p210). On one interpretation, favoured by McKittrick, this background has no reality, and should be thought of along much the same lines as matter was by the medievals: as a kind of privation or contingency which things only exhibit to finite intellects. Rosenberg

himself suggests such an interpretation in a recent paper, arguing that everything is ultimately fully actual relative to the highest level individuals which subatomic individuals help to constitute (2015, p241). But, as the previous chapter argued, this leads to mind-independent objects only having reality because natural individuals (paradigmatically, organic life-forms) are there to “complete” them, as their final causes; in other words, a kind of idealism. However, Rosenberg also seeks to draw connections between the background of possibilities and quantum indeterminacy, “in which microphysical entities are represented as sets of potentials” (2004, p210), and points (as mentioned) to the possibility that physics will one day identify an ingression with the “layered decoherence” of indeterminacy into the familiar world of solid objects (2015, p240). This would seem to grant independent reality to the background after all, and identify how it could be investigated: by physicists, not philosophers.

Montero (2001) goes so far as to wonder whether physics has left the physical behind altogether; “Current physics, which posits such things as particles with no determinate location, curved space–time, and wave–particle duality, tells us that the world is indeed more ghostly than any ghost in the machine” (62).<sup>12</sup> Brian Greene’s (2000) popular science book on string theory illustrates Montero’s point clearly. Having reduced the properties of spacetime to fields generated by the vibrations of sub-microscopic, one-dimensional ‘strings’, Greene wonders how the strings ought to be conceived independently of their relational properties. Since they are themselves supposed to be “‘shards’ of space and time” (378), strings cannot be conceived as spatiotemporal outside of the patterns they weave. Prior to these configurations,

---

<sup>12</sup> Montero notes that Russell (1927/1992) made the same point, p. 78. See also Chomsky (1988/2001), p142-6.

we can ask ourselves whether there is a raw precursor to the fabric of spacetime—a configuration of the strings of the cosmic fabric in which they have not yet coalesced into the organized form that we recognize as spacetime... in the raw state, before the strings that make up the cosmic fabric engage in the orderly, coherent vibrational dance we are discussing, there is no realization of space or time. Even our language is too coarse to handle these ideas, for, in fact, there is even no notion of before. (*ibid*)

Understanding of this strange primordial realm has just begun, and Greene gestures tentatively towards a recently developed mathematical framework known as “noncommutative geometry” (379), in which “the conventional notions of space and of distance between points melt away”. Montero comments that the new physics hardly qualifies as physicalist anymore; it seems closer to a sort of Platonism. “Speculation about such nonspatial, nontemporal stuff (or perhaps it would be better to call it ‘nonstuff’) should ... be a bit worrisome for those who define the abstract over the nonspatiotemporal— do we want to say that our spatial world emerges out of abstracta?” (Montero 64) But Platonism, in which reality is radically transcendent to appearances, was sternly criticised by Aristotle (Shields 2013, n.16, p228). And in its modern incarnation – if Montero is to be believed – special pride of place is given to physicists, not philosophers, common sense, or first-person experience, in describing transcendent reality, via the use of mathematical concepts. Rosenberg does not clearly distinguish this view from his own, in which science, philosophy and experience are on equal footing.

Moevs’ epilogue on modern physics suffers the same ambiguity (188-9), citing Heisenberg’s (1962) argument that the “smallest units of matter” are more akin to forms “which can be unambiguously spoken of only in the language of mathematics” (*Physics and Philosophy*, p160). This is more akin to Platonic than Aristotelian metaphysics, as it seems to

grant a kind of rarefied reality to matter, as well as an equally rarefied – mathematical – mode of access to it, rather than identifying reality with form and reducing matter to a calculating device like imaginary numbers. Subsequent references in Moevs' account cement this conclusion, e.g. Pagels (1982); "everything that ever existed or can exist is already potentially there in the nothingness of space" (p244). This kind of potentiality could have existed without the universe having ever come into being; precisely what Aristotle's rejection of prime matter ruled out. Aristotle may be right that no experiential meaning could be given to such a substance, which seems to exist in no particular (qualified) way. But the importance of mathematics to going beyond experiential possibilities was already recognised in Descartes' time, and, as Patterson argued, was a source of inspiration for Descartes' dispute with Aristotle.

Suppose substance were identified with the latest cutting edge physics (which is alleged to have abandoned conventional concepts of time, space and matter). Then physicists could make claims about its properties which entail massive revisions to how subjects are instantiated: for instance, by showing the only possible "layered decoherence" (as Rosenberg puts it) from possibility space to three-dimensional reality results in a multitude of subjects, none of them clearly identifiable with our conscious selves (as Lockwood argues), or a single subject that is likewise indifferent to the conscious/unconscious distinction (as Wilkes argues). Perhaps what we think of as the present moment will have to be revised in light of the latest research; this too would seem to undermine a major component of our certainty about conscious experience, that there is an immediate present.

For Rosenberg, the relation between claims made by physics, philosophy and introspection concerning prime matter is problematic given that it is the latter two which he seeks to justify, yet only the first has made substantial conceptual advances on this issue



since the time of Aristotle. Since he needs the safeguard of physics to justify his metaphysics, Rosenberg is ambiguous about the nature of the pre-existent possibility space; viewed independently of theoretical paradigms employed by physicists, it seems pretty close to unintelligible, as McKittrick complained. We are, after all, being asked to imagine something which is neither conceptual nor experiential, something structureless, timeless and contextless, and the very nature of pre-conscious 'experience' of which the background is composed is left wide open.<sup>13</sup> We have only the vaguest idea of what panexperientialism could mean, as Rosenberg admits, and what can be said of it seems to be a matter of defining it negatively over and against consciousness. It was really the metaphysics of form (subject) which came to Rosenberg's rescue, showing that the subject and sense data of introspection still coincide in us despite being ontologically separable. So in the end he is committed to Aristotle's dictum that to be is to be rationally intelligible.

Despite claiming to have radicalised Salmon's account of scientific knowledge in terms of giving an explanation via the theory of causal constraint, in doing so Rosenberg has really regressed to a deductive-nomological model instead. His argument for what is missing from Salmon's theory of causal responsibility proceeds via an analysis of the concept of causation, which in turn dictates natural laws for the composition of objects in general. For Salmon, phenomena are explained by situating them in the context of wider empirical reality, whereas Rosenberg thinks an explanation would be incomplete without tracing objects back to their non-empirical, contextless metaphysical source. Similarly, Salmon did not seek to analyse the concept of causation so much as explain it away, in terms of primitive notions that were descriptive of contingent structures in the universe, and therefore compatible with Humean scepticism about 'hidden powers' (see Dowe 1992,

---

<sup>13</sup> As we will see, this rhetorical question is posed in a slightly different context by William Wimsatt against philosophers who seek to understand consciousness in purely first-personal terms, rather than by situating it in a wider (third-personal) framework. See [3.6] and [4.5].

p200). So Rosenberg's philosophy of science is yet another example of his ambivalence between incompatible alternatives.

All of which leads to the situation where reality for Rosenberg both is and is not mental, is and is not intelligible, is and is not determinate, etc; defending the existence of a mind-independent reality that is supposed to resemble modern physics, while insisting that it can only be best understood a priori. Science's radical departures from common sense suggest that Descartes was right to be sceptical of Aristotle's comfortably intelligible worldview, in which reality is ultimately identified with appearances.

When modern naturalists seek to rehabilitate forms, they do so on the assumption that forms have been discovered inductively: they are not thought of as necessary accompaniments of all material objects, with which all of us would be directly acquainted upon turning our attention to the objects in question. Stephen Boulter goes so far as to argue that induction is the *only* method of discovering what forms there are in nature (Boulter 2007, p63), which he refers to as "non-logical" necessities. This is a decidedly modern move, yet Boulter thinks it can be traced to the early Scholastics. The reason is that he follows the 'Cartesian' interpretation of Descartes. Rather than following Patterson in seeking to salvage Descartes as a kind of proto-naturalist, Boulter thinks the Cartesian edifice must be torn down to its foundations in pre-sceptical medieval philosophy. I want to touch on this typically uncharitable reading, shared by Moevs (2005, p56), which tends to be framed as a polemic about Descartes' obsolescence. As we will see (particularly in chapter 6), reports of Descartes' death have been greatly exaggerated; in this chapter I want to emphasise that the erstwhile rebellion against him is above all an exacerbation of his

insistence on the finitude of human reason and on the independence of material possibility from form and teleology.<sup>14</sup>

### **[3.4.2] Common Sense and the Principle of Separability**

Let's recap once again. As previously discussed in detail in ch.1, Chalmers and Goff follow Galen Strawson's understanding of physicalism as committed to "the descriptive completeness of *physics* more or less as we know it".<sup>15</sup> Like Chalmers, Strawson thinks that outside of the case of consciousness, logical and natural supervenience go together. Goff goes so far as to argue that physicalism would cease to be a distinctive position otherwise,<sup>16</sup> and so would argue that agnosticism about the future of scientific discovery is too weak to rule out the kinds of non-reductionism he himself goes on to defend. He regards anti-physicalism as distinctive only insofar as it says we can know *a priori* that the scientific world-picture is incomplete. But for all these philosophers it is the nature of our access to subjective experience which motivates their criteria of true knowledge, not the state of science. Their understanding of consciousness lends itself both to epistemologies which prize certainty above all else (since the certainty of experience is a paradigm of knowledge), and to metaphysical views whereby, since (e.g.) the redness of red is absolutely determinate ("*this* shade before me certainly exists," etc), one might assume the existence of other absolutely determinate properties, which might then presuppose similarly rigid bearers of properties (e.g. point-particles), etc. They do this because the presence of conscious experience is likened to the presence of a kind of object, which in turn leads to modelling the

---

<sup>14</sup> I also hope to assuage the worry that rationalism in philosophy of science is a dead end in the light of criticisms from Salmon, or indeed that his account supports Aristotelian epistemology, as Boulter would maintain.

<sup>15</sup> Strawson 2008, p19.

<sup>16</sup> Goff (2015a), p380; Goff (2014b), p13 & 16-17; Goff (2010), p131.

nature of physical objects on the supposed determinacy of experience; even the inference from conceivability to possibility treats the philosopher as an expert on exotic kinds of modal entities.

As such, and despite challenging *mental* reductionism, Goff maintains that the universe is ultimately made up of a completely determinate number of microscopic entities described by fundamental physics, with our common-sense terms, other than those pertaining to o-facts, designating facts about clusters of microscopic entities. And he is entirely prepared to follow through on the consequences of his world-picture, when combined with Sider's criteria for truthmakers of references to organic life. If consciousness is an all-or-nothing property (we either have it or we don't), while the boundaries of organic life are vague, then whatever flicks the conscious light switch to 'on' cannot be tethered to the presence or absence of life-forms. Hence his recent defence of unrestricted panpsychism, unrestricted composition of o-properties, and then finally cosmopsychism, all of which, while highly extravagant, are also very simple ontologies. Goff is thus deeply indebted to the ideal of reductionism, which was once associated with particularly tough-minded physicalists such as Quine. But his commitment to this picture seems to be motivated less by the need to stay close to the scientific worldview, and more an article of philosophical faith, in which the intelligibility of certain metaphysical claims presupposes the *possibility* of having possession of 'all the facts' – even if this is impossible in practice.

What started as an attempt, on Hilary Putnam's part, of formulating scientific realism by reference to essential properties has become realism of a wholly metaphysical realm beyond the reach of scientific understanding, whose theological underpinnings are evident. Chalmers, despite reassuring the reader that he does not have "any religious inclinations"<sup>17</sup> frequently illustrates his talk of possible worlds by reference to how things would seem for

---

<sup>17</sup> *TCM*, p.xiv

an omniscient and omnipotent creator<sup>18</sup>; Goff often uses the same trope to argue for and against certain metaphilosophical questions of intelligibility (which will be discussed in the next chapter). For both philosophers, the ideal of mind-independent facts fixing what we refer to has become unmoored from any realistic means of verifying those same facts. In doing so Goff licenses himself to draw ever more far-fetched conclusions, ultimately departing from common sense altogether. Referring to the biblical scene of Lot's wife being turned by God into a pillar of salt, Goff writes:

The only reason a metaphysician need care about common sense is from want of anything better upon which to build metaphysics. But the orthodox property dualist has something better: a priori access to the complete nature of a certain feature of reality, i.e. consciousness. The orthodox property dualist should forget about common sense, and embrace conscious pillars of salt. (Goff 2014a, p90-91)

The same is true of Rosenberg, who draws on the critique of Chalmers' and Goff's conceptions of matter presented in the previous chapter, but retains their conception of mind, as a completely determinate object whose properties are conceptually separable from the physical, and therefore require independent explanation. He, too, ends up inventing a new cosmology to accommodate the mental. By going some way towards filling in the blanks in Goff's metaphysics – accounting for causation and non-spatiotemporal bonding relations – Rosenberg avoids the charge of mystery-mongering (or mysterianism) which could be levelled against Goff. But by construing consciousness as a special type of mind-independent entity, and claiming to ally himself with scientific realism, he still opens himself to the

---

<sup>18</sup> "It is useful to think of a logically possible world as a world that it would have been in God's power (hypothetically!) to create, had he so chosen." (*TCM* p35)

possibility of radical revisionism, which the Cartesian certainty of the mental is supposed to preclude.

Boulter argues that the Cartesian sceptic and Humean-Lewisian metaphysician are making the same mistake, which he calls the “principle of separability”, or (PS):

(PS) if one can conceive of X apart from Y then one can conclude that:

- i. *Neither can be reduced to the other* because X and Y are not identical.
- ii. *The existence of the one cannot be inferred from the existence of the other*, because either can exist without the other.
- iii. *One cannot be explained in terms of the other*, again because one can exist without the other, so *one cannot be the cause of the other*.
- iv. *Neither can be part of the mind-independent essence of the other*, because neither is included in the definition of the other. (Boulter 2011, p624, emphasis added)

This principle corrupted our trans-historical common sense, which Boulter construes as a heuristic tool for navigating our ancestral environment,<sup>19</sup> into a self-validating worldview that would pretend to explain everything on its own terms. For example, Goff characterises his anti-sceptical certainty as grasp of special, mental properties (Goff (2014b), p8-9 & 17). They are special by being separable from properties known by any other means, they can only be conceived first-personally, and so their appearance cannot be the appearance of anything not already contained in the first-person concept of the phenomenal. Against such stringent demands, Boulter argues that an adequate form of certainty is already available

---

<sup>19</sup> Heuristics are representations of the world involving simplification and schematisation, as Boulter put it above. I say more about this theory of cognition’s connection to naturalism in section [3.5].

without needing to retreat into scepticism, and which cannot be understood in the first place so long as the (PS) sets the agenda.

Boulter traces the beginnings of the (PS) back to scholastic philosophy's progressive overvaluation of logic, which then informed its metaphysics. The trend began with William of Ockham (c.1285–1347), a philosopher working soon after the 1277 Church condemnations of metaphysicians perceived to have placed limits on God's omnipotence. Since it was conventional to allow God to be limited by the principle of non-contradiction, Ockham held that only what was logically possible was truly inviolable and hence the only basis, by way of what can be imagined without contradiction, for arriving at metaphysical truth. Philosophers thereafter found themselves increasingly forced into credulity with regards to sceptical scenarios made possible by the presumption that lawlike regularities in nature were sustained, at every moment, by God's free will, and could at any time be suspended. Long before Descartes' attack on Aristotle, this had the effect of undermining Aristotle's approach to causation and ontology, in which causal acts are temporally extended, and inseparable from the propensities of actual, proximate matter. Instead, the sheer possibility of matter assuming any conceivable form took hold of metaphysicians' imaginations, and causation (or God's will) came to be conceived of as instantaneous, yet temporally discrete from, and prior to, effects, culminating in Hume's scepticism about causation (Shields 2013, p77-8).

Boulter thinks sceptical arguments against causation, direct realism, moral realism, free will, and the propositional attitudes are all question-begging appeals to the (PS). In all of these cases our access to the phenomena in question is called into doubt by their logical separability from the events, beliefs or feelings they were invoked to explain. Reconstruction of these phenomena in an ontology of Humean supervenience, as recommended by the Canberra Plan, is therefore likely to be misleading. Boulter's main point is that there are limits to the logical intelligibility of nature, and that common sense can be a default point of

reference in spite of it being justified “non-logically”<sup>20</sup>. At first blush this seems to commit him to an appeal to intuition, and one quite favourable to non-physicalist, or indeed anti-scientific, claims. For example, qualia seem to be rationally unjustifiable givens in just the same way as causation, direct realism, moral realism, free will, and the propositional attitudes are for Boulter. On the contrary, however, Boulter thinks suspending the (PS) will undermine anti-physicalists’ thought experiments, wherein one thing being causally necessitated by the other (e.g. minds by the presence of brain activity) must imply that the opposite (one thing not being necessitated by the other) is inconceivable. Boulter thus follows the standard a posteriori physicalist line of allowing that zombies are conceivable, while denying that this necessarily indicates the existence of a possible world where zombies are actual; the space of possible worlds might just be smaller than it seems to be.<sup>21</sup> But he presents this as faithful to Aristotle, for whom there existed “non-logical” necessities in nature which could only be discovered a posteriori.

By denying the propriety of seeking only those ideas which are clear and distinct to a perfectly self-conscious intellect abstracted from worldly relations, Boulter hopes to expand the meaning of conceivability to include ways of knowing which are dependent on real physical properties.<sup>22</sup> This, he believes, will recover our ordinary, common-sense certainties about how the world is – certainties which nevertheless do not stand Cartesian scrutiny – and enable new appreciation for the kinds of necessities governing the mind-independent world which he claims science now does far greater justice to than it did 300 years ago. A highly plausible way of applying Boulter’s metaphilosophy to the problem of consciousness was touched on earlier in Chapter 2’s discussion of John Heil [2.3.1], who, like Boulter, thinks

---

<sup>20</sup> He also brings an evolutionary argument to bear on why common sense ought to be presupposed until non-question begging arguments against it are supplied, but I will set that aside for now.

<sup>21</sup> Discussed in more detail in Chapter 4.

<sup>22</sup> This puts him in good company with both philosophers of science who reject deductive-nomological models, and neuroscientists who deny that knowledge is essentially propositional in form. See [3.6].



he is returning to Aristotle. But I want to suggest that no such return is possible, thanks to modern science's incompatibility with direct realism. And to the extent that Boulter and Heil take *indirect* realism on board, their positions cannot be as intuitive or commonsensical as they claim.

### **[3.4.3] John Heil and the Fate of Direct Realism**

Like Boulter, Heil is sceptical about inferences from conceivability to possibility (Heil 2015, p49), and affirms the inseparability of the mental from physical properties. As we saw in [2.3.1], there are significant overlaps between Heil's view of what this inseparability amounts to, and panpsychism – and in fact Heil credits Michael Lockwood as an inspiration (Heil 2004a, p242). For Heil, causal dispositions are simultaneously qualities, and the two can no more exist apart from one another than matter and form. It is crucial to understand, however, that this view, although superficially similar to Chalmers and Goff's metaphysics, gives Heil good reasons for departing from the philosophical strategy (which follows the Canberra Plan and presupposes Humean supervenience) by which they arrive at anti-physicalism, and for downplaying the reliability of introspection.

In Heil's ontology, neither macroscopic objects' properties, nor those of macroconscious states, are logically supervenient on a base of microproperties/qualities; a priori analysis cannot get us from one to the other, even though macro-and micro-properties are inseparable, the latter nothing over and above the former (Heil denies reality is comprised of levels). He denies that properties bear analysis: where there is a sharing or resemblance of properties between two objects, this will be a "primitive, internal" relation, a matter of resemblance "*tout court*" (Heil 2004a, p196). Heil claims this is common-sensical:

Cooks explain qualities of dishes they have prepared by citing qualities of their ingredients; painters explain the qualities of different colors and textures of paint by reference to qualities of constituent pigments; audiophiles trace the qualities of amplified sounds emitted by loudspeakers to qualities of the components. (Heil 2004b, p245)

In all of these cases, people must simply gesture towards or list the qualities contributing to the final product. Given the impossibility of analysing qualities, the capacity of a certain combination of qualitative characteristics to give rise to others cannot be predicted a priori; the possibility of blue and yellow combining to make green could not easily be predicted by someone unacquainted with green. In these and other cases – such as optical illusions – seeing is believing; they would strike us as impossible otherwise. Heil would therefore agree with Churchland that there is no special hard problem of consciousness distinct from other problems of emergence which we do not understand (see [1.2.1]).

Heil turns the intrinsic, brute nature of qualia against proponents of Cartesian certainty about our mental states: since similarities and differences between qualities cannot be analysed, we could always “err in judging a state to be of a particular sort when it is in fact a state of a different, *though similar*, sort. Is the feeling in the pit of your stomach hunger or nausea? You may find it difficult to say” (Heil 2004a, p229). Moreover, if sensory qualities cannot be analysed a priori, then we do not have a “good grip” on their “precise nature”, and so cannot know in advance whether their a posteriori analysis into qualities possessed by brains is possible (Heil 2004a, p239). Privileged or veridical access to one’s own states would then be “trivial” (Heil 2003a, p237), a point I return to in chapter 4. To be sure, Heil is still a kind of constitutive Russellian monist, but he makes good on Goff’s contention that that position is hard to distinguish from physicalism, especially where deference to neuroscience over first-person intuition is concerned. That deference to authority also

manifests in another point of divergence from Goff, in that Heil is noncommittal about the sorts of objects that exist, despite often seeming to take the middle-sized objects known to common sense as paradigms of objecthood. In fact, Heil's suspension of judgement on this score is inextricable from his commitment to objects in the first place. He recognises the close connection between properties and universals, but unlike Aristotle, he refuses the move of ultimately reducing objects to modes of the unmoved mover (a universal)'s being. Heil takes common sense for granted at least in the thought that real objects cannot just be bundles of universals. In this, Heil follows philosophy since Descartes in identifying actuality with particulars rather than universals, and affirms the existence of at least one substrate – albeit one necessarily possessing properties, rather than the formless stuff of so-called prime matter.<sup>23</sup>

Heil notes the connection between substrate and subject; for Berkeley mental substance was the only substance there was (recall Rosenberg on the need for a subject to glue qualia together). But his deference to science prevents making much of this: for all we know, he says, “there is but a single object: space, or space–time, or some all-embracing quantum field. If that were so, then ordinary objects would turn out to be modes of the one all-inclusive object. A beetroot, for instance, might be a red, spherical, pungent region of space–time” (2003a, p177). By this Heil does not wish to deny the reality of middle-sized objects: he merely says that concepts and predicates refer to vague similarities and differences rather than sharply defined properties (Heil 2003a, p57-8). But it means there can be no clear boundary between the presence or absence of consciousness in middle-sized objects like us, which is precisely the puzzle that led Goff to cosmopsychism, and Wilkes to eliminativism or behaviourism (despite her disavowals of these labels), and led to the

---

<sup>23</sup> He also denies that knowledge of an object's properties amounts to knowledge of universals, because he downplays the metaphysical relevance of similarity or sharing of properties (2004a, p193-6). Heil therefore does not have the problem of accounting for the mind's access to universals, although, as I will argue, he does have the related problem of accounting for knowledge of mind-independent qualities.

previous chapter's attempts to shore up a place for real medium-sized essences that could give ontological weight to the self.

Rather misleadingly, Heil (2003a, p211-3) compares his view of the self to that of Kant and Wittgenstein, both of whom defended a non-empirical and indubitable self (Kant's transcendental unity of apperception, which accompanies all thought, and Wittgenstein's self as "limit of the world" in the *Tractatus*), and made it central to their disputes with physicalism. For Heil, the properties these philosophers attributed to the self are simply properties of our self-representations, which could be studied scientifically. Self-centric consciousness is one way among others for representing qualities – not proof of metaphysical unity of the universe (as it was for Dante). Nor is it proof that to be is to be perceived – since qualities can exist, and be referred to, when unperceived. Heil follows Lockwood's disclosure view: everything is qualitative even when no-one is looking. Consciousness is not a matter of shining a light on mindless stuff in order to bestow qualia on it; it's just a matter of becoming aware of one's own qualitative thinghood, and its causal-perceptual relationship with other qualitative things, whose self-disclosure can be represented *as* their disclosure to a self (i.e., the body with which they have qualitative/dispositional relations via secondary qualities). Having concluded that the capacity for self-representation underlies all this, Heil goes on to claim homing missiles have "a primitive point of view" (214). So despite challenging functionalism for its claim to be able to abstract mental properties from substances (see below), Heil ends up not so far removed from Chalmers, a functionalist who is open to even thermostats having very simple experiences – since they represent their internal states to themselves. Everything hinges, then, on whether qualities can do the heavy lifting required of them in Heil's ontology. We can see their importance – but also some of the motives for dispensing with them – in comparing Heil to Wilkes.

Like Wilkes, Heil defers to cultural norms and practices, themselves a loosely defined object or 'power net', to set the boundaries of concepts such as tables. And descriptive facts about what we find interesting or important – facts probably rooted in our evolutionary history – restrict predication of these concepts (Heil 2003b, p216-7). Wilkes (1988) takes the opportunity to point out that games, tables and ornaments do not have “necessary and sufficient conditions” that could be identified *via negativa* through imagining the concepts in increasingly unfamiliar and unsuitable settings. By contrast, she says, thought experiments in physics specify the relevant background conditions whose variation could affect the outcome of the scenario, lest it be inconclusive (7), *before* setting up the idealised case study (e.g. frictionless planes). She draws a line between semantically indeterminate common sense terms and the “clarity and economy” of scientific background theories informing thought experiments. Moreover, natural kinds only serve a “useful purpose”

to pick out groups of things which science finds it useful, profitable, convenient to isolate. They ... provide, in the main, the central *explananda* and *explanantia* for systematic study: they are the terms for which, and with which, the laws and generalisations of science are framed. Hence 'water', 'mass', and 'tiger' are natural-kind terms; 'fence', 'ashtray', and 'ornament' are not. (Wilkes, 1988, p13-14, emphasis added)

In particular, despite having ruled out as unscientific (and hence inconclusive) thought experiments that cast doubt on personhood – e.g. Derek Parfit's discussion of split brains (Parfit 1971) – Wilkes nevertheless draws the conclusion that, where split personalities are involved, our concept of a person simply breaks down, and defers to science to supply “non-rational” causal explanations underlying such deviations from folk psychology. For Wilkes, it seems fairly clear that the difference between science and common sense (she discounts

philosophy as a source of insight here) has to do with the clarity and distinctness of scientific concepts; a recognisably Cartesian move. And the distinctiveness of scientific knowledge draws on its authoritative relationship with a mind-independent substance underlying the objects we are familiar with. Given the assumption that there is such a thing as matter in itself (matter which is not just form in disguise, as it was for Aristotle), it is hard to see how analysis could ground out in macroscopic qualities; substance inevitably retreats to a realm beneath that to which the senses have access, and common sense is at most second best.

For Heil, the disappearance of reality from the sensory world is supposed to be blocked by the fact that all objects possess qualities, at least some of which are manifest in experience. The strength of this view is that we already know what it would mean for an object to possess both powers and dispositions, since the only types of objects we encounter possess both. What a quality is seems hard to grasp without making essential reference to our experience of objects, the way they appear – they are paradigm cases. Even primary qualities, which persist independently of experience, are still qualities we can experience. Even if paradigmatic objects, such as beetroots, turn out to be less fundamental than we thought, this does not mean that at least one object is not present to experience. So Heil's view seems to presuppose a form of direct realism. But that is exactly what he cannot countenance, and in fact Heil explicitly distances himself from direct realism. This is because his solution to the mind-body problem must account for the observed qualities of the brain being different from those of experience. A brain has spongy, moist, grey qualities to it which it does not share with a tomato (Heil 2003a, p225-6). And yet "experiences are in the brain" (*ibid*, p223). Heil responds – and he follows Lockwood here – that "we must distinguish qualities of your visual *experience* of a tomato and not qualities of the *tomato*... it is the tomato that is spherical and red, not your experience" (2004a p225). This raises the

question, if your experience of a tomato isn't spherical and red, then what qualities does it possess?

Heil's way of putting it lends itself to being read as a defence of the topic-neutrality of sensation. However, he adds that qualia "seeming difficult to describe stems from your having learned to ignore them, your having grown accustomed to treating them as 'transparent' indicators of the qualities of perceived objects. Your description of them, then, would, unavoidably, be framed in terms of the objects of your experience" (*ibid*, p226). Perhaps, then, with practice we could begin to describe the real properties of experience without reference to the outside world. But now we seem to be losing a grip on what qualities are, or how they relate to mind-independent objects and powers. In fact, the situation would not be improved if the qualities of experience were perfectly describable, since Heil has yet to account for why the experience of a tomato should resemble a tomato, and how we could know if it did. In doing neuroscience, he says, "we are comparing an experience (of a tomato, for instance) with an experience of an experience of a tomato" (2003a p234). And those experiences will be relative to the sorts of creatures we are: Heil refers to perceptual systems as the "material of representation", which can include ad-hoc systems such as use of a stick to feel under the fridge as much as biological ones (Heil 2003a, p227). Felt qualities "are not to be identified with qualities being represented." Heil predicts, on this basis, that tactile visual stimulation systems or TVSS (used to substitute for eyes for the blind) generate qualitatively different visual sensations from eyesight by virtue of their matter, rather than by formal (functional) differences between the two.

This raises the question: are objects describable independently from the qualities by which we represent them? Conversely, if every way of qualitatively representing an object captures something about it, how robust is Heil's realism? Objective reality seems to have dropped out of the picture here – unless we fall back on a topic-neutral analysis of

experience which could accommodate its being visible as a brain, and which downplays the differences between representational materials whereby objects are represented as having such-and-such qualities.<sup>24</sup> Heil points out that a causally indirect relationship with the outside world could still be epistemically direct; such a causal relationship might just be what it means to perceive anything. Lockwood would agree (see especially Lockwood 1990, p300-1), however, his account of epistemic justification insists that we have no access to mind-independent qualities, and sharply distinguishes between these and an object's causal effects.

Heil's ontology credits Locke's distinction between primary and secondary qualities, but denies the latter can be understood merely as pure powers of the former, since they have a reality of their own (Heil 2004b p244). Lockwood follows Berkeley's attack on Locke's distinction and affirms pure powers as our only contact with things in themselves (Lockwood 1990 p151); he thinks it is incoherent to talk of analogy or resemblance between experiential properties and mind-independent properties, even "geometrical" ones (Lockwood 1990 p156). So Heil ought not to see similarity between their positions – except for the fact that Heil's indirect realism should pressure him to accept the rest of what Lockwood says. Both philosophers remain committed to things in themselves being qualitative. They would both claim that projecting the property of qualitiveness onto inscrutable matter is not just a hypothesis, but a conceptual requirement; causal realism and qualia-realism are mutually presupposing. But what kind of qualities are being attributed? Clearly they wouldn't be

---

<sup>24</sup> One problem with this approach, as far as Heil is concerned, is that it focuses on the causal role played by objects acting on perception in abstraction from the ways they can be perceived, i.e., it gives a functionalist account of perception. But Heil's emphasis on the material of representation was an attack on functionalism. For instance, he thinks thought would be extremely impoverished without imagistic imagination, but the latter makes essential reference to qualities which depend on our particular constitution (2004a, p237). Bechtel (2007) makes a similar move in his discussion of the sociology of science; "Going beyond what is known calls upon imaginative resources in which analogy, mental images, "leaps of thought," etc., can play a key role" (p34)



qualities as we know them, since the full range of experience only ever puts us in contact with qualities of the brain.

Just as Heil only proved that there must be at least one object, without showing what it must be like, so the requirement that there be qualities tells us nothing about what they would be like. But if they are nothing like our qualia, what is the point in extending our concept of qualia to them – except as a gesture of goodwill? Arguably, of course, this point applies to the concept of objecthood, as Berkeley pointed out; why hang on to the label ‘object’ if paradigmatic objects, such as beetroots, are actually modes of a bigger object which we cannot even imagine? The concept of a *real object*, for Heil, has become just a placeholder for whatever science discovers, much as c-fibres and Newtonian atomism stood in for contemporary science in Lewis’s metaphysics, and for this reason Heil’s metaphysical realism could be criticised (by e.g. Ladyman and Ross) as rather tenuous. After all, how can science be talking about qualities, as he claims, if real objects are totally cut off from experience? Indeed, in a dramatic echo of Goff, who was criticised in chapter 2 for distinguishing real essences from scientific ones, Heil writes that “physics is silent on an electron’s qualities”, studying only “quantifiable relations among objects” (Heil 2004b, p244).

Finally, if following Berkeley and putting objecthood and qualities in the same boat seems objectionable, that may be because primary qualities (or better: primary *properties*) seem more robust precisely because they are not qualitative, and therefore less relative to a point of view. Heil is appealing to resemblance between qualities of common sense and something unknown. Conceiving of reality as non-qualitative can just as well be viewed as a confession of ignorance as to what it could be, and this could be because we’re convinced perception does not put us in contact with things themselves. As Moevs argues in his exposition of Aristotle, a crucial component of the latter’s realism was in *nous*’s ability to

transcend the medium of representation in order to grasp universals; this requirement ruled out the mind's dependence on any ordinary physical stuff (Moevs 2005, p82). As he goes on to explain, being limited to the medium – being unable to transcend it – is to lose the world, since it rules out the self-transcending grasp of universals. Meanwhile for Heil, it undermines the possibility of identifying mental qualities as variations of the qualities shared by all things, since we lack a point of reference to mind-independent qualities. If we do have access to objective reality, then, it will not be qualitative. Heil says that “What it is like to experience seeing a bowl of flowers with your eyes is not what it is like to see the same bowl of flowers via a TVSS” (Heil 2003a p229). It seems reasonable to think of what those two types of representation have in common as a primary quality. But that quality will be more like what is common to Descartes' wax across its many manifestations than something strictly present to experience; as Lockwood would put it, it is a property imposed on experiential objects by reason. Our access to primary qualities, if we have one, is likely to be intellectual, not perceptual.

I conclude that Heil's ontology, according to which qualities and powers are inseparable, cannot provide what the previous section sought: a plausible explication of what Boulter's attack on the principle of separability (PS) amounts to, with reference to the mind-body problem. So what else could suspension of PS mean? Rather than appealing to paradigmatic objects of common sense in order to illustrate what it means, I suggest naturalists are better off appealing to sceptical arguments.

#### **[3.4.4] Naturalism and Scepticism**

On the face of it, denying (PS) also rules out the kind of deep correspondence between reason and nature sought by Aristotelian and medieval mystics. Indeed, the impossibility of

such a correspondence seems to be exactly what guarantees nature as a mind-independent order in the first place, in contrast to the idealistic implications of identifying nature with form (for Boulter the forms are not themselves instantiated or manifested by the intellect). And Wilkes's stated sympathies (Wilkes 1988, p43-4) for Richard Rorty, who denied we could think imperishable truths because we are perishable things, gives the lie to her own supposed Aristotelianism. At the same time, and in seemingly dramatic contrast to her Rortyan historicism, Wilkes also insists on a strong form of metaphysical realism, arguing that "few would claim that *unimaginability* entails *impossibility*. Results from spacetime physics, non-Euclidian geometry, and quantum mechanics show this conclusively" (*ibid*, p17).<sup>25</sup> This is similarly unfriendly to Aristotle, however. So there seems to be a problem here, one which similarly affects Boulter.

Boulter focuses on how the (PS) leads philosophers to overvalue the space of logical possibilities, but their doing so is really a symptom of an even deeper commitment to having a God's eye-view in which it is they, as thinkers, who are separated from everything else. Giving up on this means giving up on nature as an inherently intelligible order. While he claims to be faithful to Aristotle, Boulter's attack on the principle of separability is really the assertion of a mind-independent material world against those philosophers, including Aristotle, who wish to reduce it to *nous* or intelligibility. This is in line with Descartes' assertion of the independence of extended substance from thought; Boulter just gets rid of some theological prejudices which Descartes' account seemed to strengthen, such as that prime matter needed to be maintained by God because it lacked a vital spirit or form of its own. But it is difficult to understand how the material world, so understood, could enter into our philosophical theories if it is unintelligible to reason. Berkeleyan idealism still seems to

---

<sup>25</sup> Wilkes traces this attitude back to the beginnings of modern science; Robert Boyle writes "I see no necessity that intelligibility to a human understanding should be necessary to the truth or existence of a thing." (Boyle, 1672/1772, p.450 )

beckon. It is ironic that Boulter attributes Cartesian scepticism to overconfidence in logic – identifying being and intelligibility – when his own attacks on modal rationalism can profitably be viewed as an even deeper scepticism, and which may be closer to what Descartes had in mind anyway.

The solution is not a recourse to idealism, but instead a reassessment of naturalism's proximity to scepticism, a proximity which Boulter (though perhaps not Wilkes) is keen to disavow. Boulter criticises Hume's scepticism of causation for resting on "unforced assumptions regarding the nature of necessity and possibility" (Boulter 2007, p77), much as I have done in response to Lewisian influence on analytic philosophy of mind; but he and Wilkes both view disagreement among philosophers as a reason to suspend commitment to any of the available options (Wilkes 1988, 16-7; Boulter, 2007, p13). And this metaphilosophical move – refusing the force of an argument because it rests on presuppositions which have been, or could be, disputed – can be traced back to Pyrrhonian scepticism. Sextus Empiricus, the only ancient Pyrrhonist whose writings have survived, "describes scepticism as an ability to produce oppositions, among arguments or impressions on the same topic, in such a way as to produce suspension of judgment because of the "equal strength" (*isostheneia*) of the items opposed to one another" (Bett, "How Ethical Can an Ancient Skeptic Be?", in Machuca, (2011), p4). However, Wilkes and Boulter's aim is relevantly similar to Patterson's Descartes, contrasting the doubtfulness of philosophical and folk metaphysics to the "coercive evidence from science, forcing the adoption or rejection of any given solution" (Boulter, 2007, p64). Of course Descartes conceives science as an intellectual enterprise, whereas these later authors identify as empiricists, but I will suggest in chapters 5 and 6 that their differences can be dialectically reconciled.

In the meantime, Pyrrho must be distinguished from the "dogmatic" or "Academic" sceptics who asserted the impossibility of knowledge, with whom he is often conflated.

Instead, he refused the force of *any* positive claim, including those asserting the impossibility of knowledge. Peter S. Fosl (2011) draws a line of influence from Pyrrho to Hume, and argues their conclusions were in fact congenial to the same kind of ‘common sense’ naturalism Boulter and Wilkes wish to defend. In line with the distinction between dogmatic and non-dogmatic scepticism, Fosl distinguishes between two kinds of beliefs: those arrived at by a choice to assert one over another (the dogmatic position), and those arrived at by “mere yielding to assent rather than deciding to believe” (Fosl 2011, p158). Hence, “In the wake of the skeptical arguments Hume philosophically assents only to non-dogmatic beliefs” (*ibid*, p146). Hume himself sought distance from Pyrrho, whom he identified, scathingly, with “the repudiation of all belief” (*ibid*, p152), but Fosl regards this as a “hyperbolic and misleading caricature” (*ibid*) of Sextus’ *Outlines of Pyrrhonism*, which in fact “associates such a total epistemic rejection with Academic skepticism” (*ibid*).

A similar accusation may be levelled against Boulter’s reading of Hume. For that matter, one can find the influence of Pyrrho in what Boulter says as well. Boulter interprets the true task of philosophy as the quest to resolve “co-ordination problems”, in which claims made by different “first order disciplines” or “domains” such as science, theology or common sense, come to clash with one another (Boulter 2007, p11). Boulter’s “metaphilosophy of common sense” is to assert that “if a common sense belief clashes with a *philosophical* theory or argument, the common sense belief is always given the benefit of the doubt” (*ibid*, p21-2). In spite of situating himself among philosophers, such as Wilfrid Sellars, who seek to solve co-ordination problems philosophically (*ibid*, ch.1, n.7, p202), Boulter’s suspicion of philosophical reflection seems to do away with the business of second-order co-ordination altogether, and puts another first-order domain – common sense – in its place. In practice, Boulter’s efforts are directed towards discrediting other philosophers’

attempts to solve coordination problems by appeal to un-commonsensical conclusions.<sup>26</sup>

Boulter treats the metaphilosophy of common sense as a means of shifting the burden of proof (*ibid*, p24), but what is meant by proof is unclear; sometimes it seems to indicate empiricism as the final arbiter of co-ordination problems (since non-logical necessities are discovered a posteriori), at other times the burden could only be carried by an a priori argument whose presuppositions are themselves beyond dispute, and which necessarily commands assent – which surely sets the bar too high. Rather than a normative discipline issuing instructions on how best to release second-order reflection from aporia, Boulter’s common-sense philosophy seems to be purely descriptive: common sense will co-ordinate between domains, until it doesn’t – or until it ceases to be common sense.<sup>27</sup>

Boulter’s response to philosophical aporia seems unpromising. However, it happens to share a common thread with Pyrrho’s response to philosophical dispute, and Hume’s response to his own sceptical doubts. Unlike apocryphal accounts wherein Pyrrho’s followers had to save his life when he ignored obvious peril, such as the edges of cliffs (Fosl 2011, p152-3), a crucial part of Pyrrho’s teaching was that, having reached a state of tranquillity in which the opposing sides of all arguments have equally minimal force, one “defers to what appears to be the case” (*ibid* p156). Similarly, “Hume writes that he does not “know” – does not have knowledge of – what ought to be done, or how he can respond, but can only “observe what is commonly done” ([Hume (1978)], p268)” (*ibid* p163). Having doubted away religious and metaphysical sources of justification, he falls back on “custom, habit, and

---

<sup>26</sup> The next chapter presents a plausible approach to making good on common-sense philosophy’s promise to revolutionise co-ordination problems, with reference to the mind-body problem, though like Heil’s aforementioned attempt to do the same, it will turn out to be inadequate.

<sup>27</sup> This is in line with Boulter’s appeal to natural selection as the safeguard of common sense, since he claims our survival depends on having reliable access to truths about morality, responsibility, folk psychology, causation and perception which mostly align with folk wisdom. The sticking point is that those truths must be construed topic-neutrally, as referring to any conceivable physical realizer, and it is not clear that this is defensible. I return to this issue in chapter 4 and 6.

history" (*ibid* p161). Fosl calls this "sceptical naturalism" (*ibid* p147), whereby "nature actually completes itself in convention" (*ibid*).

What is crucial, then, is not whether or not such conventions are grounded in metaphysical essences or even in an independent, external natural order that reason or perception lay hold of. The distinguishing feature, rather, is resistance – what Sartre describes as a "coefficient of adversity" – the extent to which we find in experience that certain practices or beliefs are not plastic, revisable, or open to revision. Along just these lines, for Hume the "natural" signals not an independent causal order but, rather, the stable, the useful, the common, the easy (when we go along with it) or resistant (when we do not), the not easily revisable and regular features we recognize of the world and ourselves, and what is natural to us can at least in part be artificial. (*ibid*, p162-3)

At first this may seem to be simply pragmatism, not naturalism. But the force of habit and instinct may be taken to "mean something like the claim that our acts are really just "automatic," "instinctive," and "determined" (*ibid* p163). The Humean mind is not *resistant* to naturalisation, nor is its 'discovery' amidst sceptical doubt particularly useful – a point I return to in chapter 6.

Several questions are left outstanding at this stage. There is the question of how Humean empiricism could be reconciled with the Cartesian rationalist picture sketched by Patterson, which was what motivated belief in a mind-independent material world in the first place. For that matter, given the theological motivation for Descartes' 'geometrical method' of arriving at his idea of extended substance (since it follows the Church condemnations of 1277 in crediting only logical possibility and the principle of separability with explanatory power), it is not clear if the older Aristotelian concept of matter as privation, finitude or nonbeing was ever truly refuted. Could the two concepts of substance

be reconciled? There is also the puzzle of how deference to appearances could fail to bring us back to the problem of qualia, which seems fatal to physicalism because qualia resist naturalisation. There is the question of the status of unobservable entities, which are foreign to custom or habit, but are postulated by modern science. And conversely, theories which extend far beyond appearances must somehow be reconciled with direct realism, if the latter is, as Boulter insists, the most plausible upshot of denying the principle of separability. After all, “In Hume’s skeptical theory, our relationship to the world is ... not most basically one of knowing” (Fosl 2011, p157). Whatever Boulter means by non-logical necessities, it seems as if he means them to have rational force, and not simply the force of habit. He and Wilkes also endorse essentialism (a fairly strong form of metaphysical realism), despite both, as we have seen, harbouring views closer to scepticism. All of these questions will be addressed over the following chapters. To begin with, I want to emphasise how Hume’s sceptical naturalism is mutually supportive with naturalised psychology, whose supporters are once again keen to portray it as thoroughly anti-Cartesian.

### **[3.5] Functionalist and Teleological Semantics**

A crucial part of naturalising the mind is identifying how minds represent the external world – what it could mean in purely physical terms to perceive that something is the case, think about things, including things with no causal effects on oneself (e.g. the future), or intend to do something. An influential attempt to solve this problem is now known as the teleosemantic or teleofunctionalist approach to representation. This approach is associated with a number of philosophers, including Ruth Millikan (1984), David Papineau (1993), and William Bechtel (2007). Boulter (2007) also leans on it to set up his “metaphilosophy of



common sense". But the doctrine is descended from earlier functionalist or 'systems' approaches, and typically compared and contrasted with these. The systems approach is defended by Cummins (1975), but I will draw on an article by Robert Van Gulick (1980, reprinted in Lycan 1990), in defence of the systems approach, and the following exposition will highlight its similarities and differences with its successor, teleofunctionalism.

Van Gulick (1990, p109) poses the problem as one of how brain processes can be thought of as recognising or representing something. Clearly there is "non-random covariance" between these processes and the environment. However, for whom does this encoding of information represent anything in the world? Van Gulick sees the obvious answer as: the organism as a whole. But non-random covariance does not in itself denote the possession of information *about* features of the world. The organism must be able to *use* the information appropriately for it to be said to possess information, rather than the latter being merely 'stored' in the organism somehow (like, e.g., a barcode on the back of one's neck). Thus for information to be possessed in this way is for it to play a dispositional role – i.e. how it could influence the organism, given the right circumstances? Consequently, such information need not be 'in the head' or open to reflection, but merely capable of being adapted to, as with our response to extreme heat. In any case, the organism adapts to its environment both passively, in the manner of plastic in a mould, and actively, in the sense that it makes no sense to speak of adaptation except as a modification of existing tendencies toward some state or states. Van Gulick argues that though the attainment of muscle mass through exercise constitutes an adaptation, it isn't covariant with particular features of the environment and thus doesn't represent anything.

He then argues (111) that information implies the existence of goal-directed behaviour; the organism's goals must be enhanced by its having certain facts available to it, if it is said to possess information at all (though such goals could include simply the need to

be roughly aware of one's surroundings). Van Gulick admits that his definition of adapting to, as opposed to merely responding to, the environment on the way towards some end-point is vague. One would not want to characterise the obstacles encountered by a rock on its path down a hill in anything like the way an oncoming car could affect one's intention of crossing a road. In such cases, there are facts about one's internal structure that affect responses to the world – while internal features are wholly irrelevant to the behaviour of rocks. Van Gulick's approach suffers from what Papineau (1993) terms "disjunctivitis" (p58), in which the internal mechanisms responsible for triggering some belief in response to a stimulus could be fooled into entertaining a false belief even if they are working perfectly, in the sense that they would work the same way in cases of veridical perception (a real-world example of this would be frogs' inability to visually distinguish a fly from a bee-bee pellet, and consequently will try to consume both). Since it is the functions underlying perception that individuate it as perception, according to the functionalist picture, cases that we would *prima facie* regard as misperception would have to qualify as *accurate* representations of a disjunctive content, e.g., fly-or-bee-bee. Teleofunctionalism solves this problem by appealing to natural selection: supporting the ability to reproduce provides the norm by which to gauge a trait's success at representation. After all, "biological purposes are always a matter of *results*", and true beliefs "get the organism to *behave* in a way appropriate to certain circumstances" (Papineau 1993, p59).

Van Gulick sees information possession as a question of the complexity of "systematic interaction[s] of component structures within the system" (111) that determine behaviour. At the same time, such attributions of goal-directedness would seem to be a matter of taking stances or interpretations, as much as properties of real patterns (112). The intentional stance affords a look at abstract patterns that are not readily detectible in the microphysical causes at work in a system (there is a sense in which nothing really *acts*

according to beliefs or desires, any more than objects in motion *do* physics – rather, they *embody* our best descriptions of their behaviour).<sup>28</sup> Furthermore, this stance allows one to place a system in context (or conversely, assuming such a stance follows from examining the wider context); a membrane is more teleological than molecules performing a similar function in isolation from cellular activity (punning the word 'functioning' is useful here, he suggests).

Van Gulick later (126-7) suggests that taking stances is a matter of having limited information: were we (impossibly) in possession of all the processes involved in an agent's activities, we would not need to attribute or infer any contentful states, since all relevant causes could simply be read off from the physical facts. In contrast, Papineau (1993, p65-7), like Boulter, stresses that teleosemantics safeguards folk psychology from this sort of scepticism; “Natural selection favours things which produce certain effects. But it can't favour things which don't exist”. Macdonald and Papineau (2006) point out that the systems approach is ultimately descriptive, not normative, “it doesn't seem to show that a trait in any sense *ought* to be doing F; it just says it *isn't* doing F, and so is statistically unusual, but nothing more” (11-12). Once again, the normativity implicit in the concept of an evolutionary adaptation comes to the rescue: something that is not a matter of interpretation must answer to our concepts of beliefs, desires, and other propositional attitudes (concepts that are themselves normative). Teleosemantics therefore allows for a topic-neutral analysis and defence of common sense; in particular, a topic-neutral, anti-essentialist account of meaning.<sup>29</sup>

---

<sup>28</sup> This way of thinking has been popularised by Daniel Dennett throughout his career. The intentional stance is an “interpretation” of behaviour that seeks rationality, agency, purpose, belief/desire states, and other mental entities not strictly manifested by behaviour, in order to predict and control that behaviour. See Dennett (1991), p76-7. Other physicalist philosophers have objected to the instrumentalism implied by Dennett's approach to folk psychology (e.g. Papineau 1993, p65-7).

<sup>29</sup> Boulter extends this claim to cover many other aspects of common sense. For example, he argues that “common sense demands only that space and time be of such a nature as to allow objects to be located at a distance from each other, and for events to be sequentially ordered. It has nothing to say beyond this about

Van Gulick sees his account as extending mental holism (the interrelatedness of belief and desire) to account for what it means to intelligently possess, as opposed to merely store, information. Context affects what it means to believe or desire anything – such as other beliefs or desires, as well as facts about the world. Thus any such system must be able to recognise relations between facts and have at its disposal a wider context (e.g. a theory) in which to place occurrent facts. And this demands a level of isomorphism between functional organisation and information content – the web of beliefs must be (functionally/dispositionally) inscribed if it is to have any effects. Such a holistic account suggests that there will be no adaptive response to just one piece of information – intelligent use of information presupposes a background grasp of far more facts than one is currently paying attention to. To possess all possible facts would mean to possess a mind that was neither embedded in an environment nor in possession of an implicit grasp of relevant facts and past precedents for the significance of any piece of information. A Laplacian demon would not only have no need to adopt the intentional stance towards others, it could not interpret itself as a mind either – and nor could we, since it would be a non-physical mind of a sort totally unrelated to beings like us (and so could not be a counterexample to Papineau’s realism about folk psychology).

On the functionalist account of content, to possess information is to manifest the right behaviour in a given context, or be predisposed to doing so given the opportunity. But this is not a simple stimulus-response account of knowledge, since the ways in which information is used are determined by the background of related facts and dispositions. Meaning, rationality, and the propositional attitudes are consequences of holistic structures

---

the nature of space and time, in the same way that common sense has nothing to say about the chemical composition of water. These theoretical matters, while of the greatest scientific interest, had no ecological or social import in the ancestral environment, and are of no interest to common sense” (Boulter 2007, 43).

and dynamics, so that one can intelligibly claim they are realized in hardware of some kind. To be rational is to possess a minimum of coherence between propositions one endorses, as well as paying attention to their logical consequences, which fits with arguments that delusions are not really beliefs, since they do not fit into a network of other beliefs in such cases. (Exactly how minimal this consistency may be is not specified.) By contrast, Van Gulick argues (118-9) that frogs' tendency to flick their tongue in the direction of any fly-sized object passing in front of them is too much a case of stimulus-response to be characterised as involving beliefs (e.g. "there is a fly in front of me", or even "whatever it is, it might be food"). This fits in with his earlier remarks on the necessity of having a range of responses or attitudes available whenever information is *intelligently* taken on board. Being able to draw inferences from or report on particular pieces of information seems to suggest that intelligence must be reflexive – one must be able to introspect that one believes X, and for what reasons, in order to properly believe it. Information pertaining to e.g. muscle fibre groups is registered in the brain, but though this helps regulate behaviour, the same information cannot play a role in one's conscious assessment of muscle fitness, and cannot inform other networks of belief.

However, as Bechtel (2007, p170) emphasises, even representations have their origins in motor control. Primitive regions of the brain respond to stimulus from the muscles; this basic system of co-ordinating (representing) the motions of muscle cells then gets re-represented by higher control systems responsive to more abstract input (e.g. instructions to push a button). For this reason, the teleofunctionalist account stresses satisfaction of desires as key to caching out a naturalist account of truth and representation. To be sure, Van Gulick (120) allows that to understand something is to "have behaviour-regulating mechanisms which allow one to deal with it in a way enhancing the realization of one's goals", so to have information at all will be relative to one's "goals or interests". However, his approach to

teleology only refers to the internal economy of an organism's brain functions. For Van Gulick, "cognitive operations involve only formal operations on representations. Whatever exists beyond the representations themselves and the procedures for operating on them is irrelevant for understanding cognition" (Bechtel 2007, p178). His approach is therefore "methodologically solipsist" in a way that teleofunctionalism is not. Teleofunctionalism necessarily makes reference to the mind-independent context and history in which organisms are situated in order to individuate their mental states.<sup>30</sup>

As we have seen, according to Papineau (1993), "a desire's real satisfaction condition [is] that effect which it is the desire's biological *purpose* to produce." But we can also go further and "pick out the real truth condition of a belief as that condition which it is the biological *purpose* of the belief to be co-present with" (1993, p58). Because directing behaviour via desires is more fundamental than directing desires via beliefs, only the former can help naturalise what we mean by 'truth'. Roughly speaking, then, "The truth condition, for any belief, is that condition which guarantees that actions generated by that belief will fulfil its biological purpose of satisfying desires" (Papineau 1993, p80).

Van Gulick briefly touches on the implications this account has for scientific beliefs: he suggests that they are adaptive in certain environments, such as determining one's status in the scientific community. Alternatively such higher level information states may be said to be derivative of more basic kinds of response to one's immediate environment – having reliable responses to 'manifest perceptibles' is a prerequisite for reading scientific instruments, for example (Papinea 1993 takes this line, p75, emphasising the compositionality of beliefs). For Van Gulick, "what information we possess about such a system and in what ways we understand it will depend on how we are able to adaptively interact with it" (Van Gulick p120). Thus there is a sense in which our understanding of the

---

<sup>30</sup> This has the counterintuitive consequence of allowing for the possibility of beings with no mental states due to their having just popped into existence, i.e., having no historical context. I say more about this in chapter 6.

world is richer simply in virtue of how many ways there are of negotiating it.<sup>31</sup> For his part, Papineau insists that his analysis of truth

isn't a theory about why we should *want* truth. It's a theory of what truth *is*: namely, for a belief, the obtaining of a condition which guarantees that, if an agent *were* to act on that belief, the ensuing action would satisfy desires. This doesn't presuppose that anybody will actually act on the belief. Nor does it presuppose that the only reason for wanting the truth in respect of that belief is to be able to act so as to satisfy desires. To be sure, if you do want to satisfy desires, then [my analysis of truth] does immediately imply that you have a motive for wanting the beliefs behind it to be true. But that leaves room for other motives for wanting truth, in both practically significant beliefs and practically insignificant ones. In particular, it leaves room for truth to be valued as an end in itself. (Can't we now ask: *why* should truth be valued as an end in itself? But I take it to be a virtue of [identifying truth with the satisfaction of desires] that it allows this as a significant question.) (Papineau 1993, p74-5)

Van Gulick supports Papineau's picture when he argues that we shouldn't "draw a distinction between instrumental and non-instrumental (or purely theoretical) understanding" (1990, p121), since functionalism on his account collapses that distinction (recall the pun on 'function' earlier). Rather the contrast should be between implicit,

---

<sup>31</sup> The capacity for "decoupled representations" – higher order representations of conditions not currently present – is a crucial sign of intelligence and, at the highest levels of abstraction, self-consciousness, since with the capacity for conscious deliberation comes the ability to think about and choose from an indefinite range of things. Hence it would seem that modular brain functions are inevitably un- or sub-conscious, while what is globally available in the brain is at least a candidate for consciousness. This is not to suggest that frogs are automatons; Van Gulick denies that they are (p119) – but he argues that their intentions could not be specified in natural language without anthropomorphism (what the alternative might be is not clear, though it would presumably be non-folk-psychological. Similarly, what the precise relation might be between sentential and non-sentential knowledge is not stated, only that there is a threshold at which folk psychology becomes pertinent to describing the difference). On this point, Van Gulick and Papineau are in agreement: I return to Papineau's discussion of animal consciousness in chapter 6.

opaque, and procedural understanding (such as a bird's wings telling us something about aerodynamics), and “explicit...inferential richness characteristic of theoretical thinking” (*ibid*). This means that

the progression from the frog to the physicist [is] a move toward systems which become 'aware of' (understand or possess information about) the rationales which are implicit or embedded in their functional structures... Coming to have a more objective understanding of x is to acquire a capacity for the sorts of indirect dealings with x that result from placing it within a larger inferential context. (*ibid*)

True representations are an achievement – the result of a slow expansion of understanding the context in which things happen so as to encompass the universe at large, albeit always by approximation. The drive is towards greater abstraction, rather than strict correspondence truth; Papineau (1993, p99-101) stresses that teleosemantics is incompatible “with verificationist analyses of meaning which imply a conceptual tie between the truth conditions of judgements and the conditions under which those judgements are asserted” (*ibid* p99). This has the interesting consequence of recursively affecting Van Gulick’s status as a realist or anti-realist about folk psychology (FP), for it suggests that stances are not simply points of view one may adopt at will towards any system, but instead demonstrate (by their utility) a sophisticated mode of *access* to abstract information (one could compare this with Heil’s realism about statues as objects produced by a “power net” that includes human culture and values). Van Gulick argues that meaning is use (Van Gulick 1990, p122), along with information theory's basic claim that to be is to make a difference,



and ties this to the meaning of symbolic representations. One knows one's way around the world, and the brain is just more world.<sup>32</sup>

If functionalist or teleofunctionalist accounts are on the right track, then the way in which representations represent might be thought to undermine Chalmers' modal rationalism, and Goff's Cartesian certainty about the powers of introspection. As I noted above, to really imagine having full knowledge of the universe, or even just of oneself, would require being a non-physical kind of mind in the first place. To reiterate the message of chapter 2, anti-physicalist arguments that lean on the possibility of omniscience are therefore question-begging. But not only does biology portray the reach of thought as ultimately no greater than the reach of causal influences mediating between the brain and the universe, it may also force us to revise what we mean by external reality in the first place. To see why, I now turn to a recent attempt to rethink philosophy of science in light of naturalistic psychology.

### **[3.6] The Heuristic Psychology of Science**

Section [3.4.4] began by problematising the notion of a naturalised Aristotle, at least as he is portrayed by Wilkes and Boulter, due to their proximity to scepticism. Having gone 'through' scepticism to a naturalistic picture of psychology that Hume could have endorsed, we are in a position to understand how essentialism, if not full-blown Aristotelian metaphysics, could be salvaged. Characteristically, Boulter traces his essentialism back to the medieval scholastics:

---

<sup>32</sup> I return to this issue in the next chapter.

... while the natural order is always richer than our conceptual representations of it, our concepts can be adequate to real singulars because simplification is not falsification, and because at least some of these noticed resemblances are grounded in objective features of things in themselves, namely, their individual substantial form. (Boulter 2013b, p75)

As will become clear, Boulter's admission that conceptual representation must by necessity simplify its subject matter, and yet enjoys real causal relations with it, is importantly similar to the heuristic nature of mind-world relations uncovered by contemporary psychology. However, we must be careful. Boulter strongly suggests a relation of imperfect but more or less direct correspondence between concepts and reality. As Papineau emphasised in the previous section, such an account is ruled out by teleosemantics, as is direct realism in any epistemologically significant sense.<sup>33</sup> Wilkes, meanwhile, is more instrumentalist about natural kinds, and occasionally sounds anti-realist about science's relationship with reality.<sup>34</sup> Since these philosophers' metaphilosophy of common sense leans on teleosemantics (as discussed more below), I suggest their naturalism could only be salvaged by dropping the commitment to direct realism and correspondence truth.<sup>35</sup>

Luckily, a much milder form of realism can be aligned with the cutting edge psychology of heuristics, defended by William Wimsatt (2007) and William Bechtel (2007). Drawing on a great deal of psychological research as well as social science, these philosophers make good on Lockwood's hope of providing an account of scientific cognition which is compatible with scepticism about the reliability of the senses to accurately

---

<sup>33</sup> Though as I argued in [3.4.4], Boulter seems rather disinterested in epistemology anyway, since he regards the problem of justifying knowledge claims as a holdover from the (PS), and his common-sense alternative seems to be descriptive rather than normative.

<sup>34</sup> E.g. Wilkes (1988), p199-200, where she describes paradigm shifts in science, such as the move away from Newtonian physics in the 20<sup>th</sup> Century, as simply the replacement of one set of presuppositions by another, and her comments earlier in the book (p47-8) that there are no trans-historical conditions of intelligibility; had the world been different, we would think differently.

<sup>35</sup> Chapters 5 and 6 will seek to establish that the very idea of a naturalism compatible with common sense, and in any significant sense anti-Cartesian, is likely to be a mistake.

represent the external world (the kind of scepticism Patterson's reading of Descartes emphasised). This scepticism was pushed by Hume to its logical conclusion, but in response he offered a 'naturalistic' solution. Wimsatt and Bechtel's accounts fill out the details of what a Humean, 'instinctive' or habit-based picture of belief-formation amounts to, while also beginning to show how it may be reconciled with a certain form of Cartesian rationalism, as I will discuss in more detail in ch.5.

Heuristics are "inductive pattern detectors" (Wimsatt 2007, p10); "different sets of false assumptions [or] idealizations", which "play crucial roles in teasing apart different aspects of the causal structure of our world and permeate model-building" (*ibid*, p15). Wimsatt argues (*ibid*, p23) that the feedback mechanisms of learning would be impossible without first making mistakes to measure progress against, and error has traditionally been a generator of novelty, and thus creative solutions, in science as well. Just as Fosl emphasised in [3.4.4], there can be no concept of an external order, and no learning, without our efforts encountering resistance; learning must by necessity start with a kind of bumping one's head against the world. In keeping with pragmatism, Wimsatt would agree with Ladyman and Ross (2007, p27), who "immediately distance ourselves from the positivists" and

... align our attitude more closely with that of Peirce and pragmatism. As Putnam (1995) reminds us, both the positivists and the pragmatists sought to demarcate the scientific from the unscientific by use of verificationist principles. However, Putnam emphasizes that 'for the positivists, the whole idea was that the verification principle should *exclude* metaphysics...while for the pragmatists the idea was that it should *apply* to metaphysics, so that metaphysics might become a responsible and significant enterprise' (293, his emphasis). (*ibid*)

Ladyman and Ross (and Wimsatt 2007, p149) go on to identify doing metaphysics with seeking a unitary worldview, a project which according to teleosemantics we are, descriptively, already invested in simply by seeking to navigate everyday life. They also identify science as the most powerful tool for the extension of our native capacities for prediction and control of the environment. Wimsatt (2007, p339) likewise appeals to a ‘no miracles’ argument for scientific realism whereby its success inductively supports its methods.<sup>36</sup>

The notion that pragmatic values and metaphysical naturalism somehow coincide can be sharpened further by considering once again how physical thought is according to teleosemantics and the psychology of heuristics. Heuristics “will tend to break down in certain classes of cases and not in others, but not at random. Indeed, with an understanding of how the heuristic works, it should be possible to predict the conditions under which it will fail” (Wimsatt 2007, p76-7); notice the mechanistic imagery here. The notion of mechanism is explored by Bechtel (2007), who connects it to Salmon’s notion of science as providing explanations rather than laws (p12-13; see also Wimsatt’s approving remarks in his 2007, p171-2); “Instead of abstracting general principles and applying them to specific cases, [cognitive/neuro-scientists] focus from the beginning on the specifics of the composition and organization of a mechanism that generates a particular form of behavior” (Bechtel 2007, p4). He continues (*ibid* p13-14) that “mechanisms are identified in terms of the phenomena”, i.e. “occurrences in the world”, “for which they are responsible”; and “a mechanism consists of parts and operations”, by which parts denote “structural components of a mechanism whereas ... *operations* ... refer to processes or changes involving the parts”. Crucially, “To identify parts and operations, researchers must *decompose* the mechanism—that is, undertake a *mechanistic* decomposition—either physically as in a chemistry

---

<sup>36</sup> I seek a non-inductive argument to this conclusion in chapter 6, for reasons explained there.

experiment or conceptually.” By focusing in this way on the specifics of what is actually involved in the production of particular phenomena, scientists set themselves far more manageable tasks than the D-N model would permit; as Wimsatt puts it (p17), it is not rational to try to be perfectly rational, in the sense of occupying a view from nowhere.<sup>37</sup> In keeping with the image of thought bumping up against the world, scientists generally achieve understanding by “interventions” (35) with their objects of enquiry, which may consist of “Altering the input to the mechanism or the conditions under which it functions”, or “Going inside the mechanism itself and altering the operation of one or more of its component parts” (Bechtel 2007, p38).

Bechtel argues understanding includes ‘skilled coping’, which is understood as mechanisms of the body and nervous system interfacing in the right way with mechanisms in the world. This is crucial because if conceptual knowledge is already more physical than we thought then it could conceivably be naturalised. He gives the example (*ibid* p190) of a steam engine that could be said to ‘know’ about its pressure level via the piston which embodies or ‘encodes’ the relevant information. With this model of understanding in mind, it begins to look as if instrumental and psychological factors inevitably factor into how the scientific community models the world, in contrast to Rosenberg’s insistence that they somehow disguise what ought to really count as an essence due to the possibility of disagreement. To the contrary, for Wimsatt “the use of multiple means of determination to “triangulate” on the existence and character of a common phenomenon, object, or result” (Wimsatt 2007, p43) is essential.<sup>38</sup> This also feeds back into the model of knowledge as a

---

<sup>37</sup> From a utilitarian point of view, negative utility is fixed for goals that are impossible, so the goal of maximising utility can still be asymptotically approached to the extent that one avoids wasting too much time calculating utility. Yet how much time is necessary cannot be known in advance. The issue with factoring the avoidance of excess calculation into one’s utility calculus is that you can never actually know if you succeeded in maximising utility or not.

<sup>38</sup> He continues by noting that “as with many things, it is traceable to Aristotle, who valued having multiple explanations of a phenomenon” (*ibid*). But Wimsatt’s Aristotelian leanings are unsystematic – heuristic, even – and not evidence of a return to Scholastic epistemology as Boulter would wish to maintain.

coping strategy. As Bechtel notes (p21), scientists use many more ways of knowing their subject matter than can be put into words, using diagrams and mental simulations and an instinctive familiarity with their equipment.<sup>39</sup> (I return to this point in ch.4)

The concept of “triangulation” sheds light on Boulter’s defence of essentialism, whereby “an object’s being of a certain kind can be the ground for the claim that there is a causal connection between its non-essential properties” (2007, p95). Boulter claims we come to know the “non-logical” necessities of things, not by strict deductive reasoning, but by “the examination of an object’s “constitutive activities” over a suitably lengthy period of time” (*ibid*), a view he again ascribes to the (Aristotelian) Scholastics. Once naturalised, it becomes clear that there is no transduction of real universals from empirical particulars going on here, nor does the grasp of an essence entail knowledge of a quality, as it did for Heil. Real objects, according to Wimsatt, are ‘felt’ or implied rather than perceived; to encounter a real thing is to possess a “robust” instrumental relationship with it, which Wimsatt defines as follows:

*Things are robust if they are accessible (detectable, measurable, derivable, definable, producible, or the like) in a variety of independent ways... We feel more confident of objects, properties, relationships, and so forth that we can detect, derive, measure, or observe in a variety of independent ways because the chance that we could be simultaneously wrong in each of these ways declines with the number of independent checks we have. (195-6, emphasis original)*

---

<sup>39</sup> Wimsatt likewise writes that “with parallel independent means of support available and the net reliability of the conclusion as the only concern, there is no longer any reason to limit inferences to truth-preserving ones, and the use of good inductive, abductive, or more generally, heuristic principles may have a place in the construction of exemplary arguments-in philosophy as well as elsewhere.” (198-9)

Bechtel (2007) similarly emphasises the virtues of robustness (p34-9) on the basis that theories are underdetermined by evidence, *and* the significance of evidence (generated by interventions) is underdetermined by theory. Neither rationalism nor empiricism seem satisfactory to ground scientific understanding. Nor is coherentism a good alternative, since interesting evidence will upset a paradigm's internal consistency. It is a case of "imperfect observations of a *thing-we-know-not-what*, using experimental apparatus with *biases-we-may-not-understand*" (Wimsatt 2007, p57). Good science is therefore more tentatively understood as a process of expanding the techniques of intervention, where "the main focus is calibration, not corroboration" (Bechtel 2007, p36; Wimsatt 57-8 concurs). This is not as anti-realist as it sounds. Wimsatt points out (197) that it is part of our concept of objects for them to have a variety of properties, accessible by diverse means, in contrast to hallucinations.

Bechtel (180) describes representation as a process of filtering biologically relevant information from the raw data given to sensation. Wimsatt (181-192) similarly discusses how the borders of objects are extracted from noise. In both cases objectification (representation of the world as composed of objects) is to be regarded as, strictly, falsification: essences are themselves heuristics, just as Wilkes's instrumentalism suggested. So, *what* is it which is common to all modes of access, which is prior to our filtering mechanisms and is truly mind-independent?

Ontologically, one could take the primary working matter of the world to be causal relationships, which are connected to one another in a variety of ways... These networks should be viewed as a sort of bulk causal matter-an undifferentiated tissue of causal structures. (200)

Filtering meaningful information from the world is in large part a process of extracting boundaries from “bulk” (ibid, p181). A “descriptively simple” object will be one whose boundaries overlap no matter how it is described; e.g. a piece of granite will be divisible into roughly identical “subregions of roughly constant chemical composition and crystalline form... density... tensile strength...electrical conductivity...and thermal conductivity” (182). Bearing in mind that all possible cognition will be heuristic to some extent, Wimsatt notes that although some properties of causal networks can be treated as independent “levels” or “slices”, in cases closer to psychology and the social sciences we find conceptually irreducible complexity which baffles straightforward categorisation, which he calls “causal thickets” (ibid 200). And “if compositional ordering relations break down” in this way, “traditional formulations of materialism are inadequate for ontological reasons *because you can't say what is composed of what*, although your complex system contains nothing immaterial” (235).

The reason the thickets are formless is not due to their lack of structure or properties, like prime matter was for Aristotle; but due to an excess of properties and detail. Nevertheless, Aristotle could still complain that if we are unable to say what is composed of what, then we can form no clear idea of what exists. So this is a much thinner ontology than one might wish to associate with essentialism or scientific realism; thinner, even, than Descartes' supposedly innate concept of matter as simply extended substance, which took a nominalist rather than realist line towards essences or forms. Wimsatt is simply defining materialism against what it is not, but one can already hear Montero grumbling that doing so is a lost cause. In response, I submit that Wimsatt's definition of mind-independent nature as “causal thickets” does however fit with the earlier Aristotelian definition of matter as privation, finitude, or nothingness; only this time, these properties are to be understood as mind-independent. There is no view from nowhere under which they might vanish



completely.<sup>40</sup> The reason the thickets are knowably causal is because they present a limit case for mechanistic explanation; they are cases which mechanism can penetrate, but only inadequately, because all available means of simplifying the thickets fail. But the very diversity of inadequate modes of access to such phenomena gives a robust criterion for regarding them as real. Once again the image comes to mind of bumping one's head against the world. As Fosl put it above, our relationship to the world is not ultimately one of knowing, but of blind action and reaction: this is the picture materialism leaves us with, and with which scepticism is compatible, without dogmatically asserting that it is the case. Not only is perfect certainty and the reduction of matter to intelligible form not the first thing the knower knows (as it was for Descartes), it isn't the last either: there is no possibility of a pilgrim's journey to the survey all of reality, from whose lofty vantage point materiality and finitude would be dissolved. Thought is composed of heuristics which are themselves known heuristically. This outlook is still essentially pragmatist, but it seems strangely fitting with the pre-Cartesian metaphysics of matter as privation: thought is as it were individuated, in time and space, by the unintelligible.

### **[3.7] Relative Common Sense**

Drawing on teleosemantics' evolutionary rationale, Boulter argues that we should expect philosophy to produce aporias, especially when done on the basis of introspection: a metaphysical grasp of the world and ourselves had no particular survival value. He

---

<sup>40</sup> This amounts to denying that reality is intrinsically mental, in perhaps a more dramatic way than Montero intended when she posed the mind-body problem as a question of the fundamentality of consciousness versus unconsciousness. How consciousness may be thought to relate to reality, which I construe here as a vast, insensate field of ignorance, is discussed below in chapters 5 and 6. An analysis of linguistic reference which makes room for prime matter lacking descriptive content, against Aristotle's protestations, is provided in chapter 4.

speculates (2007, p65) that many counter-intuitive a priori claims are the consequence of cognition being employed outside its ordinary domain, which “tends to be confined to matters of practical significance” (*ibid* p43). Beyond this almost any claim can be rationalised; “If the “best of reasons” can be given for anything whatever (as long as one is clever enough) then the “best of reasons” can seduce one into error” (*ibid* p69). Reason doesn’t have a clear grip on what is the case when it operates outside self-imposed constraints, such as wishing to work within scientific institutions – and why should it, when the nature of reality itself cannot be known in advance to respect those constraints? However this is precisely why naturalists ally themselves with the sciences in the first place, because they want to avoid biting off more than they can chew. They argue that piecemeal approximation or heuristic modelling of reality is the best we can do even while we already *know* reality does not respect the boundaries we choose to draw:

As the boundaries break down ... not only is it true that others' perspectives intrude on the one you wish to argue for, but also that your perspective can seem to reach legitimately to the horizon. Paradoxically, as the perspectives weaken in their own domain, they don't retreat, like good scientific theories, but their generality appears to increase without bound.  
(Wimsatt 2007, p238)

Goff’s unrestricted panpsychism and cosmopsychism is a clear example of the kind of overreaching Wimsatt is talking about. But the same could be said for the ‘Humean’ (Lewisian) or modal rationalist metaphysical framework more generally, with its tendency toward implausible domestications of problems best left to empirical science. Boulter complains that philosophy’s tendency toward totalising explanations leads to theoretical differences which philosophers are unable to arbitrate themselves. Hence his

characterisation of science as a source of “*coercive* evidence forcing the adoption or rejection of any given solution” (Boulter 2007, p64), which might seem to suggest that science’s ability to command consensus is not entirely rational. But it can also profitably be thought of as a matter of robustness versus relatively brittle theories (Wimsatt 2007, p341). “Just as robustness is a guide for discovering trustworthy results and generalizations of theory, and distinguishing them from artefacts of particular models, it helps us to distinguish signal from noise in perception generally” (Wimsatt 2007, p57).

An example of a brittle theory would be Rosenberg’s concept of natural individuals, which predicts that the number of objects in the world would increase if a person suddenly developed a split personality, as the personalities would be the carriers of distinct causal constraints on behaviour, and receptive of quite different memories and experiences.<sup>41</sup> But as we have seen, different cultures have conceived the mind differently; for Aristotle the mysteriousness of the mind, its grasp of universals, was a property of public reality in which individuals only sporadically participated in by representing the world. Rosenberg thinks the mind is essentially private and non-representational. He thinks there is an objective answer to what receptive properties there are, in which case, in the debate between conceptions of the mind as essentially embodied and disembodied (where the latter may mean dualism or neurocentrism), only one conception can be right, which fails to tally with the heterogeneous uses of the concept of mind which Wilkes surveys in her book. Wilkes has an easier time of it because she thinks the presence of mind, like the presence of functions, is to an extent ambiguous; and Boulter’s attack on the (PS) helps supports her on this by encouraging us to think of seemingly distinct properties of mind and body as intermingled in a ‘non-logical’ way. (Wimsatt does not use this phrase, but it is nevertheless implied by his

---

<sup>41</sup> Wilkes (1988) discusses split personalities in detail, p100-167.

scepticism about the role of logically valid arguments to the way good science is practiced; see Wimsatt 2007, p32-3)

The reason all of the above may nevertheless be marshalled to rescue common sense is that, *to the extent that we know anything*, the articles of common sense will be among the things known. Common sense has instrumental value, and so tracks essential properties only insofar as these are conceived as powers capable of influencing and being influenced by prehistoric humans. It therefore takes the place of Descartes' innate ideas. Extending knowledge into counterintuitive areas does not falsify intuition. Of course this is a restriction of legitimate metaphysical claims to practical contexts, as Ladyman and Ross insisted. But the alternative way of conceiving metaphysics *does* seem, systematically, to falsify intuition. It emerged in chapters 1 and 2 that panpsychists have serious difficulty taking human beings ontologically seriously: this was Goff's complaint against Chalmers with regards to o-properties, before descending into unrestricted panpsychism himself, which undermined his initial interest in preserving our "ordinary" intuitions about personhood. Lockwood was happy to go along with Churchland's eliminativism about folk psychology, while Rosenberg turned out to be less naturalist than he claimed.

Patterson's re-evaluation of Descartes ends with a note of caution about the strength of his argument. She points out that mental and physical

substances ... are distinct because they are conceived through different concepts. Obviously this form of argument relies heavily on a correspondence between our ideas of things and the things themselves; it presupposes that things we conceive of as complete in themselves are in fact complete in themselves, in that they do not depend on anything else for their existence. In view of this, it is not surprising that the argument cannot be completed until we

have discovered that our creator is perfect and thereby affirmed the truth of what we clearly and distinctly perceive. (Patterson 2000, p98)

All of these presuppositions, particularly correspondence, were challenged by the attack on the (PS) and the teleosemantic alternative. Having rejected Heil's ontology, what could it mean to both deny logical supervenience and defend a robust physicalism? Suppose we have given up, for the time being at least, on actually reducing the mind – because it is too much of a thicket. In the absence of conceptual reduction, how do we think of physicalism as being true? After all, the mind-body problem otherwise seems like a major obstacle to Wimsatt, who insists we can know ahead of time that there's no 'magic' waiting to be discovered. The kind of naturalism counselled by Wilkes, Boulter and Wimsatt avoids making definitive claims about things in themselves, but consciousness does not transcend appearances, and so seems to resist analysis into more fundamental properties in principle, rather than as a contingent matter of practical fact. The next chapter considers a physicalist response to all of these problems.



## Chapter 4:

# The Phenomenal Concepts Strategy

### [4.1] Overview

The previous two chapters surveyed some empirically-informed reasons for doubting the deliverances of anti-physicalists' metaphysical intuitions concerning the mind-body problem, the nature of consciousness, the nature of the physical, and the nature of scientific explanation with regards to those issues. This chapter will focus on an attempt to put the aforementioned considerations to work, a new physicalist solution to the mind-body problem dubbed the 'phenomenal concepts strategy'<sup>1</sup> (henceforth PCS).

Strategists of this kind will concede to the likes of Chalmers that zombies (as well as ghosts, inverted spectra and so on) are conceivable, and that phenomenal experience lacks a primary/secondary distinction by which one might prise apart the conceivability of zombies from their actual possibility. Rather than there being two ways of conceiving of qualia, in terms of their appearance and their (purportedly physical) reality, a strategy employed by typical a posteriori physicalists, PCS theorists argue for a distinction between the types of *concepts* used to refer to physical properties. Thus phenomenal concepts do not distinguish between the appearance and reality of their referents, as Chalmers argues; but we can,

---

<sup>1</sup> Stoljar (2005).

nevertheless, distinguish between the deployment of phenomenal concepts and physical concepts to refer to the same underlying reality. If they are inclined toward a particularly strong scientific realism, PCS theorists might even go as far as claiming physical concepts likewise lack a division between primary/secondary intensions.<sup>2</sup> Physical and phenomenal concepts would therefore both present reality exactly as it is in itself, while doing so in radically different ways that prevent one from inferring the possibility of deploying phenomenal concepts from one's concepts of the physical, and vice-versa. "If this is right", argues Chalmers (2010b), "then we may not have a straightforward physical explanation of consciousness, but we have the next best thing: a physical explanation of why we find an explanatory gap" (p305). Nevertheless, opinions diverge as to how this strategy ought to be cashed out so as to avoid objections, such as that it begs the question in favour of physicalism, that it mischaracterises our phenomenal concepts, and that doing philosophy would be impossible if PCS were true.

The most widely recognised manifesto for PCS is Brian Loar's influential paper, "Phenomenal States" (1990/1997<sup>3</sup>). At its core, Loar's argument is that direct perception of mental states could preclude having all the information about them: the properties introspected could be physical, but our access to them could be such as to leave no clue of this. Physicalists can therefore apparently concede that experience is ineffable, while holding that mental properties are simply brain properties conceptualised in an ineffable way, rather than taking their ineffability as indicating the existence of a separate and non-physical set of qualitative properties left out by physicalist conceptualisations of reality. And this is due to our conceptual 'access' to phenomenal experience consisting in no more than type-demonstrative or recognitional concepts "that derive their reference from a first-person

---

<sup>2</sup> Sceptical, methodological naturalists, such as those reviewed in the previous chapter, need not go along with this of course.

<sup>3</sup> All citations refer to the rewritten 1997 publication.



perspective: '*that* type of sensation', '*that* feature of visual experience'" (Loar, p597, emphasis added).<sup>4</sup> What is referred to simply as 'this' or 'that' can be as yet unknown, without one's access to it being any less direct (since this would reintroduce an appearance/reality distinction into experience); the anti-physicalist's "mistake is the thought that a direct grasp of essence ought to be a transparent grasp" (Loar, p609). Katalin Balog (2012) writes, similarly, that anti-physicalists "are mistaken because they presuppose that epistemic gaps always indicate ontological gaps" (p18); the hope is to shift the burden of proof onto anti-physicalists to explain why our epistemic situation with regards to consciousness presupposes non-physical properties, rather than peculiar concepts of the physical (Boulter's attack on the principle of separability also singles out the same presuppositions, again with the aim of shifting the burden of proof).

In focusing on experience itself in this way we are no longer referring to objects in the outside world, but on properties of seeing itself. However this need not mean we are not referring to objects at all: 'this' or 'that' could be simply what brains look like *to themselves*, the brain's conceptually impoverished access to the second-order properties of its perceiving of the environment.<sup>5</sup> While not explicitly related to telosemantic or heuristic theories of cognition, PCS certainly seems to be compatible with them. As Loar puts it, "[i]t is hardly surprising that a recognitional conception of a physical property should discriminate it without analyzing it in scientific terms" (Loar, p602). The physicalist could well ask why brains' self-knowledge should resemble what neuroscientists know, given the expediencies of evolution, while denying that this makes self-knowledge in any way indirect. And the

---

<sup>4</sup> Another, "constitutional" version of PCS characterises phenomenal concepts as quotational, "which represent their referent as *that state:—* , where the blank space is filled by an embedded phenomenal state in a way loosely analogous to the way that a word might be embedded between quotation marks" (Chalmers 2010b, p310). I say more about this version of PCS in the next chapter.

<sup>5</sup> Despite apparently defending realism about qualia, even Heil seemed to take this line when he described perception as "diaphanous" (2003a, p227), and as "difficult, even impossible, to describe independently of the objects" they represent (228).

conceptual irreducibility of 'this' explains why we can imagine having these experiences without there existing a physical world at all, and why we can do no better than to say one kind of experience is just qualitatively different from another, without (in the final analysis) being able to say in what this difference consists (red is just 'qualitatively' different from yellow, we say: but we are at a loss as to describe *what it is* about the two qualities which explains the difference). The presence of 'this' presupposes relations neither to objects nor to any other 'thises' (my experience of red does not necessarily presuppose even the possibility of experiencing yellow). Nevertheless, it could be brain properties or functions, whose existence is parasitic on that of the wider empirical world, which *occasion* the use of phenomenal concepts (Loar, p599).<sup>6</sup> Functionalism or the identity theory could be true without any *a priori* analysis of the mental in functional or topic-neutral terms being forthcoming.<sup>7</sup>

Underscoring Chalmers' characterisation of PCS as the "next best thing" to a physicalist account of consciousness, Loar goes on to concede that, as a consequence of their irreducibility, the *existence* of phenomenal concepts cannot be physicalistically accounted for. A description of reality via physical concepts alone will not lead us to believe that other kinds of concepts could have application as well. That is, we cannot

(a)... reductively explicate the concept 'phenomenal quality' as 'property picked out by a self-directed discriminative ability', or (b) ... reductively explicate the concept 'phenomenal concept' as 'self-directed recognitional concept'. Phenomenal concepts are certain self-directed recognitional concepts. Our higher-order concept 'phenomenal concept' cannot be

---

<sup>6</sup> Carruthers and Veillet (2007) argue (p232) that PCS only opens up the possibility of physicalism being true, against critics who rule it out (see also Balog, 2012, p2). It therefore shares an important common thread with the possibility of naturalising cognition opened up at the end of Pyrrhonian scepticism, discussed in the previous chapter.

<sup>7</sup> Similarly, semantic externalism could allow for the irrefutability of solipsism from a first person perspective, this being irrelevant to the externalist analysis of meaning.

reductively explicated, any more than can our concept 'phenomenal quality'. The higher-order concept 'phenomenal concept' is as irreducibly demonstrative as phenomenal concepts themselves. (Loar, p603-4)

Loar's point is somewhat obscured by the fact that PCS seeks to explain physicalistically (essentially to 'explain away') *why* we exercise phenomenal concepts; as Loar put it above, phenomenal concepts are indeed "*certain* self-directed recognitional concepts". Put another way: Phenomenal properties are not the modes of presentation of physical properties, but phenomenal *concepts* may themselves serve as modes of presentation of the physical.

Loar's hedging stems from the fact that (1) identifying phenomenal concepts with self-directed recognitional concepts (any exercise of "*that*" in introspection) obscures the distinction between full-blown introspection, on the one hand, and blindsight, on the other (Loar, p603). As discussed in [2.4], blindsight patients are able to 'guess' what they are perceiving if prompted, and will consequently exercise a very 'thin' demonstrative (their introspection picks out a sheer 'something', at best). Short of reducing all exercises of introspection to a kind of blindsight (a charge sometimes levelled at Daniel Dennett's (1991) account of consciousness as the exercise of certain cognitive capacities<sup>8</sup>), the phenomenal concepts strategist must not seek to explain away phenomenal concepts *without remainder*. This is to say that (2), PCS does not expect us to be *satisfied* by the explanation of consciousness provided: describing phenomenal concepts as a species of self-directed cognition is not the same 'complete' explanation as a reduction of "water" to "H<sub>2</sub>O". The same intuitions about the possibility of zombies and ghosts will remain, because the identity of *this* with brain properties will still *seem* contingent. The task is then to explain why this feeling of contingency does not bring us back to the problem Kripke raised against a

---

<sup>8</sup> See e.g. *TCM*, p189-90.

posteriori physicalism, namely, that if there is no difference between the appearance and reality of consciousness, and reality is exhausted by physical concepts, then mind-brain identity should be knowable a priori. (Loar p608 is aware of this objection, but Chalmers 2010b, p183-4 accuses him of failing to meet it). Dodging Kripke's attack, on which Chalmers' metaphysics rests, is where PCS gets interesting, and develops the arguments against armchair philosophy laid out in the previous two chapters.

#### **[4.2] Strong Necessities**

As discussed in [1.1], Chalmers' main argument against physicalism is that we can conceive of the entire physical world without being compelled to conceive of conscious states along with it; because this act of imagination doesn't leave any room for hidden connections between mind and matter without physicalism being false, we can rule that physicalism is in fact false. This is intended to rule out likening the identity of mind and brain to the identity of water with H<sub>2</sub>O:

... for any *a posteriori* necessity, the terms involved have extra content in addition to their referent. 'Water' does not just refer to H<sub>2</sub>O, but expresses the property of being the colourless, odourless liquid in oceans and lakes. 'Hesperus' does not just refer to Venus, but expresses the property of being the heavenly body visible in the evening. It is these extra contents that allow us to construct 'surrogate' possibilities of dissociation alongside the necessities that water is H<sub>2</sub>O and Hesperus is Phosphorous. (Goff and Papineau 2013/4, p750)

For reasons that will become clear, Papineau glosses Chalmers' commitment in psychological, rather than metaphysical terms: The *appearance* (to us) of contingency in identifications between one thing and another depends on the presence of descriptive content getting between us and the thing itself, e.g. water's colourless, odourless, runniness (etc) contrasted to the underlying properties of dihydrogen monoxide. "By contrast [according to Chalmers], if we are thinking of something directly, without the mediation of some contingent description, then any necessary properties will be available a priori" (Papineau 2007a, p476).

For modal rationalists like Chalmers, there is a possible world for anything we can conceive of. There is one which characterises what we are thinking of when entertaining XYZ being water. Similarly, "when we entertain the hypothesis that Cicero is not Tully, this hypothesis corresponds to specific scenarios that we can elaborate" (Chalmers 2010b, p171). These scenarios, or "surrogate" possibilities, identify those worlds where the aforementioned hypotheses are actualised.<sup>9</sup> But our relation to mental properties lacks a mode of presentation that would allow them to be unveiled as a way in which physical properties appear. There is no possible world with features distinct from the actual one in which pain, as we understand it, could present itself as c-fibres firing: "If a necessarily true claim is a posteriori, then at least one of its terms must be associated with a descriptive content" (Papineau 2007a, p476). This is not to say that pain/brain relations lack intensions altogether, but there is no room in our conception of them to explain why their identity appears contingent without giving up on physicalism. For example, the possibility involved in conceiving of zombies has as its surrogate a zombie world whose properties are exhausted

---

<sup>9</sup> To clarify: reference by description is not the same as having a surrogate possibility to hand. The former guarantees that our reference is second-best to having a rigid designator, or a stable primary intension. By contrast, although Chalmers thinks there is a surrogate possibility corresponding to the thought that the mind might not have been physical, this doesn't commit him to our knowledge of mental appearances being arrived at by way of description.

by all that could be scientifically known, allowing for a distinction to be drawn between the scientific worldview and the real nature of things in themselves in order to explain why zombies strike us as impossible. (As discussed in chapter 1, Chalmers thinks the essence of things referred to indirectly under physicalist descriptions will include qualia, whose *phenomenal* properties are invisible to 3<sup>rd</sup> person scientific description.)

PCS replies to the above by claiming that identity could also be 'brute', otherwise known as a strong necessity, rather than amenable to being cashed out in terms of epistemically or metaphysically necessary possible worlds. Here the primary intension need not contain information about its referent in the form of some descriptive content: reference could be direct, such as, "arguably", in the case of "proper names, demonstratives, and observational terms" (Goff and Papineau 2013/4, p751). It seems, for instance, that we can forego descriptives such as 'greatest Roman orator' when identifying Cicero with Tully; "After all, we are surely perfectly capable of directly disbelieving truths, even when we don't muddle them up with other claims" (Papineau 2007a, p486). In these kinds of cases we are ignorant not because some deceptive mode of presentation gets in the way, but just ignorant full stop:

Why shouldn't some terms simply refer directly to their referent, and stop at that? Call a term or a concept 'radically opaque' if and only if it does not reveal any substantive information about its referent. If there are two distinct but co-referring radically opaque terms, then putting them together with an identity sign between them would give rise to a strong necessity. Perhaps 'Cicero is Tully' is like this: its falsity is conceivable—because there is not enough a priori accessible information about 'Cicero' and 'Tully' to know without empirical investigation whether or not they are identical—but there is also not enough extra

content to produce a 'surrogate possibility' that Cicero is not Tully. (Goff and Papineau 2013/4, p750)

Strong necessities are such that, simply by virtue of their obtaining, it *appears* as if they obtain, without our being able to specify *in what way* things would have to be, in order for them to count as states of affairs in which said necessity obtains. A strong necessity does not appear in any particular way: *that* it obtains is all one can say about it (thus mirroring the demonstrative character of phenomenal concepts).<sup>10</sup> Put differently, "the property that is the referent also serves as a mode of presentation" (Chalmers 2010b, p309). Hence if the cause of phenomenal concepts is physical then phenomenal concepts necessarily denote physical properties; the two are not separable (again, this recalls Boulter's attack on (PS): his non-logical necessities may be thought of as strong necessities). Qualitative similarity or difference between one's epistemic situation in a world wherein a strong necessity obtains and one where it does not is a non-issue; by stipulation there can be no possible world lacking a strong necessity (this being, once again, part of its strength):

...consider a philosophical view on which it is metaphysically necessary that an omniscient being (e.g., God) exists but on which it is not a priori that such a being exists. Then, according to this view, 'An omniscient being exists' (or *O*) is an a posteriori necessity ... If *O* were an ordinary a posteriori necessity... there would be a metaphysically possible world verifying  $\sim O$ . However, on the philosophical view in question, there is no such world. 'There is an omniscient being' does not seem to have any difference in its primary and secondary intensions, *so if a world satisfies O, it verifies O*. One could put the matter by saying that there is an epistemically possible *scenario* verifying  $\sim O$  but no metaphysically possible *world* verifying  $\sim O$ ." (Chalmers 2010b, p167-8, emphasis added)

---

<sup>10</sup> Recall that one of Aristotle's arguments against formless matter – ultimately an argument against materialism itself – was that nothing can appear in no particular way.

Papineau and Goff dub this view “modal dualism”, which can be thought of as a counterpart to the dualism of phenomenal and physical concepts. Just as both sets of concepts conceive of their referents directly, without an intervening sense disguising its referent, so conceiving of brute identities does so without any extra descriptive content that would verify them, in this or any other possible world. And just as there is no information that would allow one to infer the possibility of physical concepts from phenomenal ones, and vice-versa, so there is no way of inferring, a priori, what epistemically possible scenario corresponds to a metaphysically possible world, and vice-versa.

All of which amounts to a rather complex way of denying the inference from conceivability to possibility which remained intact even in Kripke’s exposition of a posteriori identity claims. But is direct conception of a property really, as Loar (p609) insists, compatible with this being an opaque rather than (as anti-physicalists claim) a transparent conception of its subject matter? And why does the identity between (e.g.) pain and c-fibres firing seem less plausible than the alternative, rather than both seeming equally plausible, like the identity or difference between Cicero and Tully? Surely, the objection goes, there is more to phenomenal properties than the brute fact that they appear – a claim about them that seems to be just a variation on the Identity Theory’s attempt to analyse experience in topic-neutral terms. And how can physicalists insist on modal dualism while still claiming to do metaphysics – since doing so surely requires having transparent access to modal facts?

### **[4.3] Objections**



The objections raised by anti-physicalists towards PCS all orbit around the idea that it is incompatible with a rich and substantive grasp of phenomenal and modal facts; these objections tend to support one another, and a full rejoinder to them all in tandem will be considered afterwards.

#### **[4.3.1] Dual Revelation**

PCS promises to leave our intuitions intact: the aim is to avoid “weird positions” (Loar, p609), such as epiphenomenalism or eliminativism. Drawing on Thomas Nagel’s (1974) observation that to be a conscious entity is for there to be something it is like to be it, Loar allows that “[b]eing experienced like that is essential to the property Mary conceives. She conceives it directly” (Loar, p599). But direct conception in terms of what-it’s-likeness (i.e., qualia) is touted as compatible with the essence of experience being describable by physical science under quite different concepts, e.g. functional kinds appropriate to neuroscience. One way of understanding what Loar is saying is to take him as defending “dual revelation” (Goff 2015b, p138). On a dual revelation account, both physical and mental concepts present their referents directly, rather than under a contingent mode of presentation associated with descriptive content, as in the case of the difference between Venus’s manifestation as blue at dawn and red at dusk. Dual revelation seems to have advantages. The alternative, treating phenomenal concepts as indirect representations, leads to the familiar and implausible claim that the appearance of consciousness can be distinguished from its reality, whereas treating physical concepts as indirect leads to Russellian monism or some other form of scientific anti-realism (Chalmers 2010b, p149-152).

Granting the plausibility of scientific realism, the question then is how to reconcile it with realism about phenomenal qualities. Goff argues that doing so requires finding a space within physicalism for “real acquaintance” with experience, which has a number of features:

*Revelation* – A psychologically normal subject can come to know the real nature of one of her phenomenal qualities by attending to that quality.

*Phenomenal Certainty* – A psychologically normal subject is able to put herself into a situation in which, with respect to one of her phenomenal qualities, she is justified in being certain that that quality is instantiated (where to be certain that P is roughly to believe with a credence of 1 that P). (Goff 2015b, p124)

*Phenomenal Insight* – We have rich a priori knowledge concerning our phenomenal qualities. (*Ibid* p128)

Goff also refers somewhat loosely to “rational certainty”, whereby a subject can “rule out any scenario where p fails to be the case” (*ibid* p130-1); for example, “we cannot conceive of a situation where phenomenal red is more similar to phenomenal green than it is to phenomenal orange” (*ibid* p129). In keeping with his modal rationalism, Goff seems to view both phenomenal certainty and insight as giving us *rational* certainty, and, transitively, that doubting rational certainty would cast doubt on *phenomenal* certainty and insight.

This seems like a substantial body of knowledge yielded by introspection. Why, then, should it not be obvious a priori that phenomenal properties are identical with brain properties? Chalmers takes the same line (2010b, p183-4) when he argues that our epistemic situation, according to PCS, gives us no compelling reason to assume that phenomenal and physical concepts corefer, and so no reason to prefer physicalist monism

over dualism. If anything, dualism strikes us as the more obvious inference given the total lack of resemblance between (e.g.) pain and c-fibres firing. Positing an identity between the two looks ad hoc; “there is a strong intuition that there’s some cheating going on” (Goff 2015b, p140). This leads to the more significant complaint that PCS’s handling of identity claims (its commitment to the existence of brute identities), backed up by modal dualism (commitment to strong necessities), is irresponsible.

#### **[4.3.2] The Possibility of Metaphysics**

Part of what makes PCS distinctive is the accusation of question-begging launched at anti-physicalists; “[i]t is as though antiphysicalist intuitions rest on a resemblance theory of mental representation, as though we conclude from the lack of resemblance in our phenomenal and physical-functional conceptions a lack of sameness in the properties to which they refer” (Loar, p605). This may itself seem to beg the question against Kripke’s line of argument against a posteriori physicalism, which points out that the way things seem is precisely the issue where consciousness is concerned. So PCS attacks Kripke’s formulation of a posteriori truth claims. David Papineau (2007a) summarises the latter as follows:

(I) If a necessarily true claim is a posteriori, then at least one of its terms must be associated with a descriptive content. (p476)

(II) If a necessary truth still seems contingent after it is believed, then it must have some descriptive content. (*ibid*, p480)

Where no such contingent description is to be found – where the object of reference is conceived directly – then (1) *all* its necessary properties should be laid bare to us; the object in question is then said to be conceived ‘transparently’. Likewise, (2) when we conceive of something transparently – what Chalmers refers to as *ideal* conception<sup>11</sup> - then we should not be in the dark about any of its properties, or any identity claims associated with it, and should be absolutely unable to doubt that (e.g.) Cicero is Tully.

If the phenomenal concept strategist’s modal dualism is accepted, however, then (I) is denied: there is “a primitive notion of metaphysical modality which cannot be analysed in terms of conceivability” (Goff and Papineau 2013/4, p752), meaning that there are conceivable scenarios (such as the separability of mental and physical properties) corresponding to no possible world. A conception could thus be both direct and opaque. PC strategists can remain agnostic with regards to (II), where transparent/ideal conception of all the relevant facts would guarantee we are no longer puzzled by the mind-body problem, because while we can plausibly be said to directly conceive of mind and matter, it is not clear that we have transparent knowledge of them, in the sense of knowing all there is to know about them. In particular, it may turn out that the faculties by which we conceive of things under phenomenal and physical concepts do not overlap; inability to unify the two conceptions and thus dispel the mystery could just be hardwired,<sup>12</sup> and it is not even clear

---

<sup>11</sup> As discussed in Chapter 1, Chalmers allows for both negative conceivability, in which we are unable to rule out a possibility given rational reflection, and positive conceivability, where we are “able to form some sort of clear and distinct conception of a situation in which the hypothesis is true” (2010b, p144). While the conceivability of ghosts and zombies is merely *prima facie* negatively conceivable (anti-physicalists will certainly welcome a priori arguments for their inconceivability on ideal reflection), the case for a real distinction between mental and physical properties is ultimately grounded in phenomenal properties already being ideally positively conceivable for us even in spite of our limited powers of reasoning. Hence, since we cannot be wrong about the presence of phenomenal properties, because we conceive them directly, we cannot be wrong about their difference from physical properties; or so it has been claimed prior to PCS.

<sup>12</sup> David Lewis’s “Ability Hypothesis” (Lewis 2004) likewise argues along these lines (pp97-102), construing first-hand experience of sensations as akin to abilities; know-how being both ineffable and non-overlapping with discursive knowledge furnished by the sciences. Paul Churchland, while taking Lewis’s knowledge/ability distinction as exemplifying “the crude divisions of our prescientific idioms” (“Knowing Qualia”, in Churchland and Churchland 1998, p145), takes a similar line, informed by neuroscientific evidence. And Hutto and Myin (2013) defend an updated version of the ability hypothesis, likewise drawing on cognitive science, which they

that a being without some form of incapacity with respect to its self-understanding is compatible with the plausible assumption (granted by Chalmers) that cognition needs to be physically instantiated somehow.<sup>13</sup> In any case, settling the issue one way or another would seem to be a matter for empirical (neuro-) science to discover, which undermines the autonomy of a priori moves from conceivability to possibility and shifts the debate decisively in the physicalism's favour.

Much as the aforementioned allows PCS to avoid the burden of giving a satisfying (transparent) conception of mind-brain identity, so too can it avoid explaining why we find their disunity more intuitive:

Maybe it lies in some deep architectural feature of the mind, and so will indefinitely remain a source of cognitive illusion...On the other hand, maybe it is a relatively shallow phenomenon, due to nothing more than the unfamiliarity of genuine physicalism, and so will disappear once this view ceases to seem so strange. (Papineau, 2007a, p493)

But now it begins to look as if PCS really is cheating: strong necessities are being introduced to head off any objection to the strategy, backed up by hand-waving in the direction of future neuroscience. And this is exactly how Chalmers responds (2010b p170, 181-2), firstly arguing that the postulation of strong necessities is question-begging (they are introduced solely in order to save physicalism):

---

connect with PCS's approach to defusing the mind-body problem (pp169-178). It should be added that all of these authors remain sceptical of the idea that knowledge of qualia amounts to knowledge of distinctive first-person facts, as anti-physicalists tend to argue, and that it is granting this point that distinguishes PCS from the ability hypothesis (PCS grants that introspection yields facts about experience being like "*that*"). However, there is enough overlap between the two positions to illustrate that PCS's appeal to separate faculties is not as unmotivated by the empirical evidence as it may seem.

<sup>13</sup> See discussion of Gunderson in [5.4].

As things stand... [PCS] require[s] that physical properties have their phenomenal mode of presentation noncontingently. But this means that the explanation is building in a necessary connection between physical and phenomenal properties from the start and so is assuming strong necessities in order to explain strong necessities. (*ibid* p184)

Moreover, strong necessities are irresponsible, licensing psychologistic explanations of our intuitions not just with regards to facts about mental objects, but logical and mathematical ones (*ibid* p182), which would then undermine the rationality of science insofar as it is dependent on these disciplines. If, as I argued in the previous chapter, there is a connection between Cartesian metaphysics and physicalism, it is no surprise that PCS takes an *internalist* line with regards to modal intuitions. By contrast, Chalmers maintains that the kinds of distinctions we find it intelligible to draw (between good and bad kinds of reasoning, and between diverse kinds of properties) are determined by the way things are in themselves, not by facts about us. Teleosemantics failed to take a coherent position on this issue, as it both treated mental representations as finite, contingent coping strategies located in the brain, therefore licensing psychologism, while also identifying representations *as* representations via their extrinsic causal relations to real, mind-independent objects, as externalists such as Chalmers maintain.

Goff is even more blunt, pointing out that the *ad hoc* introduction of strong necessities licences silly ontological claims such as the existence of “the spirit of the woodland”, which is strongly identical with the properties of tree bark (2015b, p139-40). This is probably a response to Papineau, who accuses anti-physicalists of committing the “antipathetic fallacy”, and consequently suggests, unwisely, that he is committing the pathetic fallacy.

Ruskin coined the phrase 'pathetic fallacy' for the poetic figure of speech which attributes human feelings to nature ('the deep and gloomy wood', 'the shady sadness of a vale'). I am currently discussing a converse fallacy, where we refuse to recognize that conscious feelings inhere in certain parts of nature, namely the brains of conscious beings. (Papineau 1993, p116)

PCS defenders have tended to be more cautious, treating mind-brain identity as perhaps the only strong necessity, but it is easy to see the temptation of solving other problems in philosophy using their methods.<sup>14</sup> After all, PCS denies there can be a transparent third-person description of 'all the facts' that could limit postulation of strong necessities to just the issues physicalists want them to solve. Yet the plausibility of physicalism over other, inflated, ontologies is limited if "the world as it is in and of itself is in certain respects unintelligible", Goff 2015b, p139). Which is just another way of saying that there is no empirical evidence motivating belief in modal dualism in the first place: cognitive science's discovery of non-overlapping introspective and discursive faculties does not support the conjecture, as Papineau concedes (2007a, p488-9), arguing that if things were that simple then we should be able to combine the two faculties in the same way we already combine our distinct faculties of sight and sound. So doubling down, as Papineau does, on the appeal

---

<sup>14</sup> e.g. Boulter's claim that causal powers are strongly necessary, and therefore compatible with the ideal negative conceivability of there being no such thing; he thus takes himself to have solved Hume's problem of induction:

"...while it can be agreed that we do not have impressions of events *having to happen* as they do, it is just as important to recognise that we do not have impressions of things *just happening to happen* either. Our sense impressions are blind to modal facts of any sort, including contingency. ... Running Hume's argument in reverse reveals that appeals to the phenomenology of impressions cannot decide the matter. Whether beliefs concerning modal facts are justified is a question that will have to be decided on other grounds, for the phenomenology of impressions provides no decisive leverage one way or the other." (Boulter 2007, p82)

Boulter also later argues along similar lines in order to identify moral properties with natural properties (*ibid*, p177-197); ironically, Goff ranks this move alongside identifying bark with the 'spirit of the woodland' as one of the sophistries permitted by brute identity claims (2015b, p140). Having given a sympathetic reconstruction of Boulter's position in the previous chapter, I should add that the strength of his argument against Hume lies in the suggestion that causal explanation must go beyond appearances (the "other grounds" on which "beliefs about modal facts" are to be decided), not in the strong identity of appearances with causal powers. I develop this idea further in the next chapter.

to strong necessities to explain away why mind-brain identity still appears contingent even to professed physicalists, who should be used to thinking of their phenomenal and physical concepts as grasping the same underlying reality, only forestalls the question of why we should be modal dualists in the first place. If anything, the fact that we can coherently think of our cognition as divided between non-overlapping concepts referring to the same mind-independent reality supports the opposite conclusion, Chalmers' 'modal rationalism', where there is one space of possibilities accessible to rational reflection, and which permits inferences from conceivability to possibility.

#### **[4.3.3] Translucent Consciousness**

There is another sense in which PCS may be said to be question-begging. It is able to reject the orthodox Kripkean inference from the indivisibility of primary/secondary qualities in experience to their irreducibility to modes in which physical properties present themselves, because it rejects the idea that experience has descriptive content. Only this can explain why the appearance of contingency is supposed to be senseless. But experience seems to be richer than that (i.e. "that"-concepts).

In his 2010 paper, "Where's the Beef?", Robert Schroer sides with Joseph Levine's characterization of phenomenal knowledge, in *Purple Haze* (2001), as being too "thick" to be described either in topic-neutral terms or as blindly pointing, without attendant description, to unknown physical properties (Levine 2001, p82-86). Instead, he proposes a hybrid of both – experience as comprised of both a system of introspectable relations with descriptive content, and partly of demonstratives, the latter designating the mind-independent aspects of mind not available to introspection. Goff calls this a "translucent" account of



consciousness, and, evidently views it as a serious challenge to his own position, having addressed Schroer's argument in several papers.

Schroer seeks to account for "qualitative similarities, not....causal relations, between experiences" (Schroer, 2010, p511). For instance:

Phenomenal purple is more similar to phenomenal red and to phenomenal blue than it is to phenomenal green. But if our introspective grasp of phenomenal purple were limited to the meagre demonstrative characterization of 'that quality again', it's not clear how we could grasp, via introspection, facts about what it does and does not resemble. (*ibid* p510)

Furthermore, "These resemblances between the phenomenal colours are not brute. Instead, they are underpinned by their saturation, their lightness, and various aspects of their hue" (*ibid* p514). Schroer's analysis only deals with colour concepts, but he is confident that similar claims could be made on behalf of other sensory modalities (Churchland 1985 attempts a structural analysis of musical perception).

Schroer thus aims to accommodate the phenomenal insight and rational certainty which Goff was earlier cited as listing among the essential properties of introspection which physicalism must explain. However, he also seeks to drive a wedge between the macrophenomenal consciousness which Goff charged Chalmers of failing to explain (so-called 'o-properties'), and the intrinsic properties of experience, our direct acquaintance with which was supposed to refute physicalism. Having reduced the intuitive properties of colour perception to relations, Schroer suggests the remainder can be explained away along the lines suggested by PCS: "With respect to its lightness and saturation", a phenomenal concept might be thought of as denoting "the quality with such-and-such level of that

element (lightness) and such-and-such level of that element (saturation)’” (Schroer 2010, p516-7). However

It’s not introspectively obvious that our introspective grasp of the simple component elements of phenomenal colour is substantial. For what it’s worth, when I attempt to learn about a simple component element of phenomenal colour (say, for example, warmth) via introspection, about all I can say about it is that it seems to come in degrees (*ibid* p520).

Goff glosses this translucent conception of consciousness “as a composite of two ‘sub-concepts’, one transparent and one opaque. I call the transparent sub-concept the ‘window’ of the whole concept, and the opaque sub-concept the ‘screen’ of the whole concept.” (Goff 2014a, p79) He compares it to conceiving of “a sphere roughly the same size as the Earth”, where “the concept of being a sphere is the window of the whole concept, whilst the concept being the same mass as the Earth is its screen” (*ibid*).

In other words, Schroer argues that knowledge of consciousness is beefy *insofar as*, and not in spite of, the fact that experience includes relational properties; it is these that are transparent to us in introspection. He embraces the phenomenal bonding problem Goff diagnoses in Chalmers: higher level mental facts cannot be inferred from their constituent parts, and vice-versa. In which case there is no need to postulate that the mental is truly fundamental; it could be made up of elements whose properties are unknown to us due to the way they are conceived, and so need not be understood along panpsychist lines.<sup>15</sup> As Goff was earlier quoted as saying, there is a wide margin between intrinsic properties being “weird and wacky” and being full-blown conscious properties alien to physical science.

---

<sup>15</sup> As I shall argue in the next chapters, there is room for these properties to be conceived along pan*pro*psychist or neutral monist lines.

The real strength of Schroer's account is in forcing traditional defenders of qualia to locate the intrinsic properties of experience as lying 'within' the colours with which we are familiar. We all know what we mean by the 'redness of red', or 'red in itself'; but it isn't clear that anybody knows what they mean by the 'saturation of saturation', or saturation in itself, independently of its relation to hue, lightness, etc. And even if we could get clearer about what we mean by these terms, in doing so we would probably only end up conceiving experience under a more detailed set of relational properties, thus pushing the problem back.

Goff could, at this point, bite the bullet and argue that we do know *a priori* what we mean by 'hueness', etc. Instead, he concedes that "we can know a priori certain facts about the internal structure of conscious states, but not the intrinsic nature of the basic elements in that structure," but goes on to argue that, even if 'consciousness' denotes purely physical or functional properties, we can still substitute "'consciousness\*', defined as 'that aspect of consciousness we understand the nature of a priori'" (Goff 2014a, p80), and - if we are panpsychists - assert that it is this that is ubiquitous throughout nature.<sup>16</sup> Chalmers takes a similar tack:

It could be that, strictly speaking, physicalism will be true of *consciousness* because *P* necessitates *Q*, but physicalism will be false of properties closely associated with consciousness, namely those associated with the primary intension of *Q*. We might think of this sort of view as one on which phenomenal properties are physical properties that have nonphysical properties as modes of presentation. (Chalmers 2010b, p153)

---

<sup>16</sup> Goff develops this thought in his book (2017a), where he considers that microphysics, and hence microphenomenal consciousness, may be less fundamental than the whole of which everything is composed. So the thought that microphenomenal properties may actually be physical properties, conceived opaquely, is irrelevant if the bonding relation by which they enter into macrophenomenal consciousness is inexplicable for the physicalist.

Chalmers' response chimes with Levine's misgivings (Levine 2001, p86) about the way PCS handles phenomenal properties. It seeks to concede that they are conceived directly, and even (on some variations) being directly incorporated into the concepts by which we refer to them. But PCS provides no gloss on how the brain could instantiate such strange concepts, any more than it can accommodate the qualitative properties which Chalmers thinks plausibly explain the peculiarity of our phenomenal concepts.<sup>17</sup> Goff's response, meanwhile (and in-keeping with his critique of Chalmers' constitutive panpsychism), highlights a sense in which the *relational* properties internal to experience itself are sufficient to motivate anti-physicalism, without needing to consider what consciousness in itself happens to be: in analysing experience into relations, Schroer forgets that the relations of which it is composed are non physical/mathematical in nature, because we conceive the whole bundle of our experiences under a kind of phenomenal concept (Goff's o-experience) as well.<sup>18</sup>

#### [4.3.4] Raffman Qualia

Building on Levine's complaint that PCS provides no explanation of how phenomenal concepts could be physically instantiated, Diana Raffman (1995) argues that empirical evidence actually rules them out: the human perceptual system is such that there can be no such things as phenomenal concepts, at least not the kind that do the work philosophers want them to do, namely, explaining the appearance of contingency in mind/brain identity claims. This is due to the discovery of slippery perceptual objects that Thomas Metzinger dubs "Raffman qualia" (Metzinger 2003, p74).

---

<sup>17</sup> As Balog (2011a) concedes, p32.

<sup>18</sup> Nevertheless, and in spite of the weight Goff puts on o-properties, it is still unclear whether we really have distinctive phenomenal concepts for them the way we do for individual token experiences. See chapter 5.

Reviewing some of the past 50 years of research in perception and psychophysics,<sup>19</sup> Raffman writes that “with rare exceptions, discrimination along perceptual dimensions surpasses identification” (Raffman 1995, p294), particularly type-identification. Colour shades, and sonic pitch, admit of extremely fine discrimination to us “in a context of pairwise comparison” (*ibid* p295), but we happen to lack the perceptual memory to reidentify any particular fine-grained property, such as Red<sub>32</sub>, or recognise it as an instance of the *concept* of that particular shade. Consequently,

because we cannot recognize determinate shades as such, ostension is our only means of communicating our knowledge of them...[as well as] communicating the *contents* of our perceptions of determinate shades. If I want to convey to you the content of my perception of a certain shade, I must present you with an instance of that shade. You must have the experience yourself. (*ibid*, p296, emphasis added)

This may seem to confirm PCS’s characterisation of phenomenal concepts as purely demonstrative. However, the latter were construed not in the ordinary sense of pointing to public objects, but as the brain’s way of representing itself. Phenomenal concepts were supposed to describe how we introspect, not how we communicate our experiences to others; recall that PCS ultimately sought to explain away qualia-talk as the upshot of our self-scanning mechanisms. As such, Loar’s speculations sound like a theory about perception which it should be possible to test; indeed his naturalist credentials depend on it. When we try to make neuroscientific sense of PCS, however, we find that

---

<sup>19</sup> Most especially: Halsey and Chapanis (1951), Siegel and Siegel (1977a and 1977b), Burns and Ward (1982), Hurvich (1981).

Although Loar typically characterizes his introspective concepts as demonstratives, they are in fact mentalese predicates — 'phenomenal recognitional concepts', he calls them. These terms presumably enter introspection indexed, as for example 'FEELS LIKE THIS<sub>1</sub>', 'FEELS LIKE THIS<sub>2</sub>', and so forth, so as to reflect their differential contents. (Raffman 1995, p298)

It may be objected here that Raffman is attributing views to Loar that he does not hold. After all, he argues early in his paper (1997, p597) that it is precisely the first-person demonstrative aspect of phenomenal concepts that grants them their cognitive significance (they are formed “from one’s own case”), so use of ‘THAT’ should not commit us to predication of properties about which we could be mistaken. In fact, it is not clear how we could ever be wrong about what exactly we are referring to if all we mean by ‘THAT’ is whatever we happen to introspect at a given time (Papineau thinks this is one of PCS’s strengths; see below). But interpreting PCS along these lines opens it to the aforementioned charge that demonstratives are too thin: how could we ever tell each instance of them apart from the others? So it is perhaps fortuitous after all that Loar is not entirely explicit about what PCS is committed to – developing its different lines of argument has been taken up by his commentators.

Granting Loar’s claim that each ‘FEELS LIKE THIS<sub>n</sub>’ has a “distinctive cognitive content”, it seems as if the ‘beef’ by which putative qualia are conceptually distinguished from one another will consist in a type-identification.<sup>20</sup> Since experience is supposed to be conceived directly, lacking the higher-order reference-fixing property along the lines of H<sub>2</sub>O’s ‘wet stuff found in lakes, rivers, etc’, we might surmise that the cognitive content in question will be non-linguistic (“mentalese”), and so easily mistaken as non-conceptual or non-

---

<sup>20</sup> Raffman also considers the possibility of merely “comparative or discriminatory” representations (1995, p306), but dismisses the possibility insofar as we only discriminate Raffman qualia as being more or less similar to one another when they occur, and have no disposition to represent, e.g. Red<sub>32</sub>, as ‘more red’ than Red<sub>31</sub> when it is perceived in isolation.

physical (for discussion of PCS's accommodation of genuinely non-conceptual content, see ch.5). Reference will then be fixed by the causal connection between concept and property, which is supposed to be strongly necessary, guaranteeing that each 'THIS' picks out a physical property. But Red<sub>32</sub> does not trigger a distinctive cognitive content, and has no direct causal connection to the low-resolution concept 'red', because resemblance to other shades of red serves as the mediating higher-order reference-fixing property so far as first-person concept-formation is concerned (Raffman 1995, p306). In other words, things only 'seem to seem' a certain way to our self-scanning mechanisms, which lack the fine grain to identify the way things really seem to us, when we perceive (as we will see in due course, in spite of seeming to defend the existence of non-conceptual qualia, Raffman argues that the *real* reference-fixing properties are only picked out by physical-functional concepts). And if phenomenal concepts can be subject to the appearance/reality distinction which qualia are supposed to lack, then our belief in qualia cannot be accommodated by PCS.

Raffman writes that PCS depends upon a "distinction among different ways of knowing or representing physical facts. One of these ways of knowing, namely the first-person introspective way, is supposed to capture how I myself represent my own first-order perceptual states" (*ibid* p300). But the point is that we do *not* 'know' about Raffman qualia at all; we have no cognitive way of accessing them. And philosophers cannot reply that experience is no more determinate than the concepts by which it is introspected, because we can succeed in re-identifying a few determinate shades of Raffman qualia. We can say 'there it is again' with respect to (e.g.) Blue<sub>1</sub>, without this picking out any particularly significant shade of blue so far as the objective spectrum of colours is concerned; that memory favours the re-identification of certain specific shades seems to be a quirk of our physiology (*ibid* p302). As such, Thomas Metzinger concludes that Raffman qualia sound the death-knell for armchair physicalism:

We cannot...achieve any epistemic progress with regard to this most subtle level of phenomenal nuances, by persistently extending the classic strategy of analytical philosophy into the domain of mental states, stubbornly claiming that basically there must be some form of linguistic content as well, and even analyzing phenomenal content *itself* as if it were a type of conceptual or syntactically structured content—for instance, as if the subjective states in question were brought about by predications or demonstrations directed to a first-order perceptual state from the first-person perspective. (Metzinger 2003, p70)

But now it begins to look as if Raffman's argument, if it is any good, should be equally bad news for anti-physicalists as it is for PCS. Metzinger, for instance, thinks Raffman qualia inspire scepticism about the accuracy of phenomenal judgements: "In terms of David Chalmers's "dancing qualia" argument (Chalmers [TCM, 266]) one might say that dancing qualia may well be impossible, but "slightly wiggling" color qualia may present a nomological possibility" (Metzinger 2003, p80). With respect to neutral monism, things are not so simple, but I will suggest that the outcome is still unfavourable in certain respects for the anti-physicalist side of the debate – so long as the crux of the mind-body problem is thought to be qualia.

#### **[4.4] Types of Conceivability**

In his famous paper, "What Is It Like to Be a Bat?" (1974), Thomas Nagel distinguished between three types of imagination: perceptual, symbolic, and sympathetic:



I shall not try to say how symbolic imagination works, but part of what happens in the other two cases is this. To imagine something perceptually, we put ourselves in a conscious state resembling the state we would be in if we perceived it. To imagine something sympathetically, we put ourselves in a conscious state resembling the thing itself. (Nagel 1974, n.11)

Recall Goff's statement, quoted at the end of section [1.2.1], in defence of negative conceivability, which was supposed to divert anti-physicalist arguments away from unpromising avenues such as trying to conceive of having possession of all the physical facts. He argued there that it is difficult for positive conception to find a middle way between merely propositional thoughts, potentially inclusive of impossibilities, and "sensory imagination", exclusive of much that we would want to include (four-dimensional objects or the infinity of space).<sup>21</sup> I would like to suggest that these are roughly equivalent to Nagel's symbolic and perceptual kinds of imagination, respectively. Granted, Nagel would probably wish to exclude conceiving of logical contradictions from strictly 'symbolic' (mathematical?) thought.<sup>22</sup> But propositional and symbolic imagination both consist in representing their objects indirectly, via (formal or informal) symbols, or non-phenomenal concepts. Presumably one could not imagine symbolically what could not be represented by concepts or symbols, and qualia seem a reasonable candidate for that. Meanwhile sensory imagination, which can represent qualia, cannot extend to what we cannot perceive, and sympathetic imagination cannot depict experiences we could never have, such as being a bat.

Sympathetic imagination is particularly limited; in his discussion of the bonding problem, Goff came to the conclusion that a full solution to it would also solve the problem

---

<sup>21</sup> Goff (2009b), p295.

<sup>22</sup> If Chalmers is right about PCS then its version of symbolic imagination will include logical impossibilities, which would be a *reductio ad absurdum* for it.

of other minds (Goff 2015a, p379), which would be an ideal kind of sympathetic imagination. Goff is pessimistic as to whether the latter kind of knowledge is available to us, just as Nagel doubts we will ever know what it is like to be a bat: the true nature of phenomenal bonding is, as Goff puts it, 'noumenal'. So the kind of imagination that would be most appropriate to positively conceiving of mental properties is, in an important respect, unavailable to us. It seems plausible to suppose that our physical concepts are exhausted by symbolic and sensory imagination, whereas phenomenal concepts are exhausted by sensory and sympathetic imagination.<sup>23</sup> Whether the physical or the phenomenal themselves are fully captured by only these types of imagination is another issue. If they are, then even without a full solution to the mind-body problem it ought to be possible to rule out physicalism. But because our powers of imagination are limited, Goff argued that we are better off restricting ourselves to negative conceivability, where the relevant states of affairs are those we cannot rule out.

It is arguable that the two types of imagination, sensory and symbolic, must be kept apart. Doing so was crucial to Patterson's interpretation of Descartes, whose "dualism derives not from the possibility of mind existing without body, but from the claim that thought and extension are attributes of distinct substances" (Patterson 2000, p77). Descartes allowed that he couldn't rule out his mind's dependence upon things that are known indirectly, such as the body.<sup>24</sup> Having concluded, at the end of the doubting procedure, that he exists, he writes that "I must be on my guard against carelessly taking something else to be this I...may it not perhaps be the case that these very things which I am supposing to be nothing, because they are unknown to me, are in reality identical with the I

---

<sup>23</sup> Sensory imagination ascribes experience to oneself; sympathetic imagination ascribes it to others. Besides the necessity of perceiving (e.g.) colours before being able to sympathetically imagine others perceiving them, sensory experience is also a necessary indicator of other minds, without which sympathy would never be relevant in the first place. An ideal sympathy would allow us to imagine being any creature just from perceiving them, which would then bridge the explanatory gap.

<sup>24</sup> Descartes, excerpts from *Meditations II & IV* (1996), reprinted in Gertler and Shapiro (2007), p166.

of which I am aware?”<sup>25</sup> As Patterson argues, “Now he knows that he exists, he must ‘be on guard’...against simply re-adopting his old prejudices. His method for avoiding this is to review his original beliefs about what he is, and to discard any which can be weakened by doubts” (Patterson 2000, p89). This is not to say that Descartes was uncertain about the introspective evidence for his existence, only that his nature needn’t have been exhausted by those properties evidenced in introspection (i.e., “thinking”). He indicates that he is aware that his conclusions are to some extent contingent on his method of enquiry into what there is. At best, he knows he can shift the burden of proof over to the Aristotelians and show that they couldn’t rule out dualism, while bolstering his case by showing his geometrical philosophy as more compatible with Newtonian physics than Scholasticism.

Descartes places no faith in the reliability of imagination: the claim that imagination is a reliable guide to metaphysical possibility by virtue of lacking an alternative is a much weaker route to dualism than Descartes’ argument, which appeals to “pure understanding”, or “intellection”.<sup>26</sup> Imagination, for Descartes, is comparable to sensory imagination defined above, as a kind of seeing with the “mind’s eye” which involves a “peculiar effort”, and is limited in its ability to conjure up details of an image in a way that the pure understanding, which simply grasps the concept of a thing, is not. On the one hand, we have clear and distinct ideas of the mental, as a thinking thing, and the physical, as substance extended in space. But the concept of the mental is discovered by reflections on the act of thinking itself, and is especially resistant to doubt, because it can be rendered as a tautology: to doubt one’s existence is to think, and for there to be thought there must be one who thinks.

---

<sup>25</sup> *Ibid*, 165. Goff discusses the same passage in Goff (2014b), p8, but reads it as a rhetorical question. Patterson, by contrast, sees the possibility of identifying the mind with material things of which one is unaware as a legitimate one, ruled out by Descartes methodological choice to identify substances with clear and distinct ideas. I follow Patterson’s reading here.

<sup>26</sup> Descartes (1984), *Sixth Meditation*, p.50-1.

Mental substance is therefore the purest object that could be thought by the pure understanding, being the precondition for all other acts of intellection.

Only after this discovery could Descartes turn his attention to the essential properties of objects with which he did not happen to be identical, whereupon his basis for giving acts of imagination the benefit of the doubt ultimately lay in his belief that, though it would be within God's power to deceive us with regards to logical conceivability, as well as natural possibility, doing so would have been contrary to His benevolence. In keeping with the thought that Descartes' system is contingent on a prior choice of methodology, this indicates that the ideal of clear and distinct knowledge is reached by an act of will; one can be blameworthy for failing to aspire to it, in the sense of inviting God's indifference as to whether one is deceived or not.

In a crucial paragraph<sup>27</sup>, Descartes writes that "*if the I is understood strictly as we have been taking it*", that is, as the self-referential certainty of knowing that one knows; "then it is quite certain that knowledge of it does not depend on things of whose existence I am as yet unaware; so it cannot depend on any of the things which I invent in my imagination". From this perspective, attempting to ground the mind in material reality would look like a leap of imagination (Patterson 2000, p91). He is, in fact, critical of the attempt to draw *any* firm conclusions from the power of imagination.<sup>28</sup> While much of Descartes' argument appears to be premised on imaginative thought experiments, he is better read as drawing a distinction between imagination furnishing us with positive, metaphysical knowledge - something he denies is possible - and as simply helping to draw out the consequences of an *absence* of knowledge about the origins of our experiences. In Descartes' sense of the term, imagination constructs perceivable possibilities, whereas the thought that one is thinking (the discovery of the *cogito*) *intuits* a conceptual necessity. The

---

<sup>27</sup> *Ibid*, 166, emphasis added.

<sup>28</sup> Excerpts from Meditations II & IV, reprinted in Gertler and Shapiro (2007), p166

paragraph continues: “it would indeed be a case of fictitious invention if I used my imagination to establish that I was something or other; for imagining is simply contemplating the shape or image of a corporeal thing.” He rejects the folk belief that the soul is “some thin vapour which permeates the limbs”,<sup>29</sup> which he associates with Aristotle. Contemplating non-corporeal things does not involve the exercise of imagination, nor does “the study of physics [,] because it employs an intellectual conception of body” (Patterson 2000, p109). Descartes writes that “our imagination is tightly and narrowly limited, while our mind [intellection] has hardly any limits; there are very few things, even corporeal things, which we can imagine, even though we are capable of conceiving them”.<sup>30</sup>

In consequence of all this, Goff seems to have been on firm ground following Descartes in prioritising a negative act of (symbolic) imagination (“how much can I doubt?”) over a positive one (“what can I imagine?”), which was his original basis for preferring ghosts to Chalmers’ zombies. That train of thought is continued in his conclusion in a recent paper that we can doubt the physical, but not the mental: a ghost is “a thing with mentality not realized in *any* stuff” (Goff 2014b, p7).<sup>31</sup> By ‘realization’ he has in mind the idea that o-properties are not in themselves fundamental, but rather supervene on something. He has used Cartesian doubt to refute constitutive panpsychism (Goff 2010, p127-8), because it claims that o-properties’ identity with the brain is inscrutable from the side of the intrinsic properties revealed in conceiving of a ghost, thereby portraying the supervenience upon micro- or proto-phenomenal properties as an extra, a posteriori fact about the essence of mind. And of course this would commit the constitutive panpsychist to denying there is a

---

<sup>29</sup> *Ibid* 164

<sup>30</sup> Descartes (1991), p186.

<sup>31</sup> As [1.4] noted, this paper was written around the time Goff was beginning to change his mind about issues such as emergence and the limits of a priori reasoning, with subsequent papers turning the absence of ‘stuff’ from one’s ghost into a positive insight into the non-mathematico-causal/structural cosmopsyché that grounds the material universe. Which is to say, ghosts need not be preferable to zombies just because the thought experiment which establishes them has fewer metaphysical commitments, as that is only an important issue if one is not keen on metaphysics anyway.

coincidence of primary and secondary intensions in our conceptions of consciousness, thus reducing something intelligible (o-properties) to something unintelligible (microphenomena).

Given all that has been said in the last two chapters, however, one might wonder whether negative conceivability can achieve all of Goff's goals without raising further questions, such as: why is the very *possibility* of doubt to be viewed as problematic, and why is the complete absence of doubt a precondition for doing metaphysics? In the context of the ghost argument, the conclusion would seem to be idealism, not (property) dualism: The physical but not the mental can be doubted, so why believe in the physical at all?<sup>32</sup> Another issue is that in privileging negative conceivability, Goff has given symbolic imagination the upper hand, because it already conceives of objects indirectly, as either coherent or incoherent, without making any claims as to whether they are really possible. Sensory and sympathetic imagination are, from this point of view, just inadequate forms of symbolic imagination, ruling out many scenarios (e.g. 4-dimensional objects and bat consciousness) that we know are possible, just as Descartes argued. The only reason why the negative conceivability of Cartesian scepticism via sensory imagination is important is because symbolic imagination cannot furnish a reason to rule out our being deceived about the existence of consciousness. But then, why *can* we rule out being deceived about that? If any explanation will do, then PCS's psychological explanation is available. PCS denies that our conception of mind includes the 'stuff' of which it is made, meaning that our primary and secondary intensions of mental and physical properties could all coincide without this fact being at all obvious. PCS thereby avoids the burden of proof of establishing that ghosts are

---

<sup>32</sup> As Papineau (1993) notes, p173-4.

incoherent<sup>33</sup>; it can concede that o-properties are conceptually irreducible, but deny that what phenomenal concepts refer to is anything but physical.

As we have seen, Goff's response has been to question the propriety of comparing mind-brain identity to cases like 'Cicero = Tully' in which there is no way of ruling a state of affairs in or out. Like Schroer, he asks, 'where's the beef?', appealing to the *positive* conception of the difference between mental beef (i.e., qualia) and the properties of empirical brains. This was the point of his appeals to revelation, phenomenal certainty, and phenomenal insight. (Goff clarifies in his (2014b) that this inference commits him to property dualism, rather than substance dualism, but the difference between the two is not important given how far Goff deviates from the scientific worldview.<sup>34</sup>) In short, the only response to PCS is to appeal to positive rather than negative conceivability.

It is true that, in general, Goff still follows Descartes in not permitting that which cannot be clearly imagined to influence how we currently ought to think about ourselves, which is why he takes conceptions of 'all' physical facts to be dialectically weak. It is also why he now appeals to simple insights, such as familiarity with pain or with primary colours, experiences which are unlikely to harbour any hidden complexity or incoherence which might only be brought to light following empirical testing. His concession to future neuroscience on the issue of zombies may consequently be read as a caveat that his arguments may one day be shown to be logically flawed. But conversely, his inference from

---

<sup>33</sup> See Goff 2014b, p17: "The physicalist might continue to insist that the conception we reach at the end of [Cartesian doubt] is in some way obscure, confused or incoherent. But she is obliged to show this, and until she does, we are entitled to suppose that it is clear and distinct, as indeed it seems to be."

<sup>34</sup> On the face of it, is unclear how phenomenal revelation could reveal both that the mental is fundamental, a property ungrounded in anything, *and* a property of something else. Perhaps the tension can be resolved by supposing that, because he thinks of physics as supervening on non-physico-causally bonded mental reality rather than o-properties supervening on the physical, Goff is arguing for a kind of one-sided dualism. This would mean that physical properties (*qua* their appearance in empirical science as mathematico-causal-relational properties) can be distinguished from qualia, but phenomenal properties, which from the side of 'noumenal' reality include unknowable bonding relations, are indistinguishable from the physical, since mathematico-causal-relational properties are *mere appearances* – recall the earlier complaint that Goff is really a scientific anti-realist. In order to motivate this anti-realism, he must treat anything ideally conceivable as at least on an epistemological par with hard-won scientific discoveries, and this is easier to achieve by appeals to positive rather than negative conceivability.

clear and distinct ideas, to phenomenal revelation, through to the discovery of qualia as essential properties, seems not to permit a materialist rejoinder at all. This might explain his confidence in a priori metaphysics, in spite of its being so at odds with the sceptical doubt by which ghosts are discovered: he now maintains that strong necessities, far from refuting the Cartesian insistence on having clear and distinct ideas of things, simply demonstrate an alternative, empirically-grounded method of arriving at them: a “stress-free modal dualism”.<sup>35</sup> Having fixed the essence of water as H<sub>2</sub>O during the course of scientific research, for instance, enables us to get clear on whether ‘watery stuff’ could ever have referred to some other chemical (Putnam’s ‘XYZ’<sup>36</sup>). Consequently, Goff concludes that “metaphysical possibility is just a more determinate form of conceivability”,<sup>37</sup> namely, “ideal conceivability”, wherein “you completely understand what you’re conceiving of”.<sup>38</sup> And simple conscious states such as colour qualia are ideally conceivable (see also n.33 above).

It needs to be repeated, however, that Goff’s criteria of clarity and distinctness are more inclusive than Descartes’. Though both Goff and Descartes defend the negative conceivability of ghosts, their inferences from this to a positive truth (the failure of physicalism) are quite different. At times, Goff seems to come out as a fallibilist:

---

<sup>35</sup> Goff and Papineau (2013/4), p752-6.

<sup>36</sup> Goff (2014b), p10.

<sup>37</sup> Goff and Papineau (2013/4), p754. Chalmers’ objection to a distinct space for metaphysical facts stems from his commitment to the sum of all facts being in principle intuitable and non-arbitrary. And Goff’s appeal to transparent concepts is intended to establish this. So, in fact he is really a modal monist like Chalmers: “both Chalmers-style modal monism and stress-free modal dualism reduce possibility to conceivability” (*ibid*).

<sup>38</sup> Goff (2017b), p289. Between them, “completely understanding what you’re conceiving of”, and “failing to find any contradiction in one’s conceptions” roughly correspond, respectively, to Chalmers’ definitions in his (2010b) of ideal (positive and negative) conceivability, and *prima facie* as well as ideal negative conceivability. Strictly speaking, ideal conceivability requires that there be no possible world in which some state of affairs obtaining is *a priori* ruled out. Chalmers’ examples of ideal positive conceivability include having possession of all possible scientific facts; ideal negative conceivability entails having addressed every possible objection to some thesis and finding it to withstand criticism; *prima facie* conceivability delimits what is conceivable “on initial consideration” (Chalmers 2010b, p143), which seems to function as a shifting of the burden of proof over to one’s opponent.



There is no empirical information I could learn which would make my conception more complete....[of course] it is difficult ever to be completely certain that what one is conceiving of is ideally conceivable... However, if something is fully conceivable, and our best efforts of reflection do not reveal any incoherence, I think this is very good grounds for concluding that what we are conceiving of is ideally conceivable. (Goff 2012, p745)

Contrast this with Descartes' criteria of absolute (because tautological) certainty about the existence of the 'I think'. The appeal of limiting oneself to negative conceivability was at least in the spirit of Cartesian stringency, since phenomenal experience's resistance to sceptical doubt imposed "an extra layer of dialectical force" (Goff 2010, p135) when compared to the conceivability of zombie worlds. But now responding to PCS has required the claim that introspection accesses certain vivid properties which are positively *unlike* those of the physical, rather than simply harder to imagine doing without (Loar allows that the nature of phenomenal concepts explains why experience seems harder to doubt). Goff's appeals to positive conceivability are much more 'sensory', much closer to a relation with a contingent empirical phenomenon; we know he does not seek to ground *knowledge* of the world in knowledge of the self, after all. And compare Goff with the following quote by W.D. Hart, who explicitly likens the discovery of the cogito to a kind of sense perception:

'...[A]s perception is our favoured, and perhaps only, basic epistemic access to actuality, so imagination is our favoured, and perhaps only, basic epistemic access to nonfactual possibility. When one has looked long and hard, when it has always looked to be the case that *p*, and when one has no good reason to think otherwise, then one has good reason to think *p*'.<sup>39</sup>

---

<sup>39</sup> Hart (2007), p122.

Hart, and sometimes Goff, think possibility can be inferred from conceivability by *examining* the object of conception under the mind's eye. Thought experiments in which mind and world are separated are simply homing in on mental properties which we were aware of all along. The object of conception – the mind – is construed as a positive fact, which cannot be reduced to an act of definition. In which case our introspective faculties could be less than perfect, since the certainty we are looking for is in the existence of the object itself. Of course Descartes also took himself to be making a great discovery, but one borne out of vigilance against the possibility that there are no objects to be certain about. Both Hart and Goff think their fallibilist approach is faithful to Descartes. Goff identifies ideal conception with 'clear and distinct ideas' (Goff 2014b, p9), concluding, as does Hart, that these put us in touch with essential properties. In support of this he cites Descartes' reply to a critic,<sup>40</sup> in which he argues that "I have never thought that anything more is required to reveal a substance than its various attributes; thus the more attributes of a given substance we know, the more perfectly we understand its nature."<sup>41</sup>

Goff nowadays seems to equivocate between positive and negative forms of conceivability, running together the conception of mathematical objects and phenomenal properties under the banner of transparent conceivability, and no longer distinguishes the assessment of a state of affairs' coherence from having a positive conception of it. For example, he refers to a million-sided object (Goff 2017b, p289) as something which we can transparently conceive of, in the sense that we know a priori that such a thing is possible. In this he portrays himself as Cartesian, but it is a more liberal understanding of what can clearly and distinctly be imagined than Descartes had in mind when he denied he could conceive of a thousand-sided object (Descartes 1984, p50).

---

<sup>40</sup> Descartes (1984), Fifth set of replies, p249.

<sup>41</sup> Ironically, something of this definition carries over to Wimsatt's definition of robustness.

In the previous two chapters we saw reason to doubt the Cartesian ‘geometrical method’ of arriving at facts about the way things in reality must be. Science does not traffic in truths indubitable to rational reflection; its methods are fallible, and its subject matter often ambiguous (another motive for thinking of it as dealing with “non-logical necessities”); essential properties are not discovered by doubting away their relations with other things, but by examining the relations into which they are disposed to enter. The decline of rationalism in science motivates doubts about the Cartesian ‘I’, doubts which Descartes himself was perfectly able to entertain. Ultimately, only the goodness of God guaranteed both that Descartes was having veridical sense-perceptions and that his standards of logical reasoning were correct (Patterson 2000, p98); once a fallibilist naturalist approach is taken instead, then the division between thought and extension becomes dubitable.

But the similarity between Cartesian foundationalism and mathematical exactness is a red herring anyway, so naturalists need not necessarily seek to downplay the importance of the latter to in order to refute the former. 17<sup>th</sup> century mathematics could give the impression of being about sensory as well as logical necessities because empirical space was thought to be exclusively described by Euclidian geometry; experience itself seems to testify that parallel lines never meet. This ambiguity as to what separates mathematical conceivability from the sensory could therefore later motivate neo-Cartesians such as Goff to read Descartes as lumping mathematical intuition in with knowledge of consciousness, and to do so themselves, in spite of the fact that for Descartes the *cogito* was more certain than ordinary logical truths, and in spite of the fact that the intellectual landscape has since changed. As the discovery of non-Euclidian geometry shows, logical conceivability nowadays has nothing to do with what we perceive in reality, but (as the use of non-Euclidian geometry in contemporary physics shows) it does seem to extend our knowledge of the empirical in ways that do not rely exclusively on our powers of observation.

The close relation between mathematics and the sciences points to another discrepancy in Goff's turn to positive conceivability. Arguably symbolic imagination conceives mathematical objects directly, because their properties are exhausted by their logical consistency. Goff plays up the similarity between transparently conceiving mathematical objects, and transparently conceiving of phenomenal properties. But symbolic imagination only yields transparent concepts insofar as they do *not* refer to phenomenal properties. And this particular use of imagination has proved to be highly reliable, in marked contrast to sensory and sympathetic imagination.

So Hart's empiricism, quoted above, rests on an ambiguity: perception might be our only access to actuality (symbolic imagination being a guide to the possible), but much that is actual is not directly perceivable. Goff's appeal to rational reflection is similarly guilty: mathematicians seem to just be doing something very different, exercising a different kind of concepts perhaps, to what we do when we engage in introspection. Having started with appeals to logical conceivability, Goff ends up really seeming to appeal to sensory imagination. Ironically, the weaker source of knowledge for Descartes has, for Goff, taken the place of the intellect as a source of knowledge about the kinds of properties that clearly and distinctly exist. And now, far from motivating the kind of scholasticism Boulter complained about in Chapter 3, it begins to look as if the distinction between the kinds of imagination, and indeed that the privileging of logical imagination in order to map the physical world actually plays into the PCS defender's hands.

#### **[4.5] The Use-Mention Distinction**

David Papineau (1993) agrees with Chalmers that experience is in some sense the ‘realisation’ of physical systems (e.g. I know I’m not a fictional entity because I’m having experiences). “The claim is only that this is nothing different from what it is to *be* a physical system of the relevant kind” (1993, p106).<sup>42</sup> Papineau goes on to point out that the challenge isn’t simply that we don’t have bat-experiences, but that if we did we would seem to *know* something – what it is like to be a bat – that is invisible from the physicalist perspective (e.g. Mary the bat-scientist would seem to learn something from being turned into a bat for a day). He responds (*ibid*, p109-10) that it is Mary’s imaginative or re-creative powers that have changed, not her acquaintance with special properties. He notes that this has the advantage of explaining how remembered experience can resemble instances of X when no X is present. “Mary’s *thinking* about the experience of seeing red would change, but what she was thinking *about* would be exactly the same thing as she used to think about when she was a scientist who had never herself seen red” (*ibid*, p111). Papineau continues:

The fact that we do not *have* certain experiences when we think third-person thoughts does not mean that we are not *referring* to them. To make this move is to succumb to a species of the use-mention confusion: we slide from (a) third-person thoughts, unlike first-person thoughts, do not *use* (secondary versions) of conscious experiences to portray conscious experiences to (b) third-person thoughts, unlike first-person thoughts, do not *mention* conscious experiences. There is no reason, however, why third-person thought about experiences, like nearly all other thoughts about anything, should not succeed in referring to items they do not use. (*ibid*, p116)

---

<sup>42</sup> Papineau (2007b, p.131) goes further and identifies phenomenal properties with the *intrinsic* properties of the physical, which he explicitly links with Russellian monism. I say more about Papineau’s flirtations with panpsychism in chapter 6, but suffice to say he is not quite on the same page as Chalmers, despite voicing approval of Chalmers’ position (*ibid*).

The use-mention distinction figures heavily in Loar's statement of PCS as well. Loar writes that "recognitional concepts are perspectival" (1997, p601), in the sense that one's point of view affects *what* concepts are deployed. "Suppose you see certain creatures up close and form a recognitional concept - 'those creatures<sub>1</sub>'; and suppose you see others at a distance, not being able to tell that they are of the same kind (even when they are), and form another recognitional concept - 'those creatures<sub>2</sub>'. These concepts will be a priori independent" (*ibid*). As exercises of sympathetic imagination, "third-person ascriptions of phenomenal qualities are projective ascriptions of what one has grasped in one's own case: 'she has an experience of that type'" (*ibid*, p597). But imputing phenomenal states to others does not necessarily *use* phenomenal concepts, as this would entail that one was literally sharing their feelings. This common-sensical solution to the problem of other minds then leads to a more important point about scientific objectivity:

Does a fully objective description of reality not still leave something out, viz. the subjective conceptions? This is a play on 'leave something out'. A complete objective description leaves out subjective conceptions, not because it cannot fully characterize the properties they discriminate or fully account for the concepts themselves as psychological states but simply because it does not employ them. (Loar, p610)

This is in contrast to Heil, for whom the use-mention distinction supported mind-independent qualities. As I argued in chapter 3, the coherence of objects possessing qualities diminishes if we think perception does not put us in contact with objects themselves. Since we do not observe qualitative objects, but only their effects on our nervous system, we cannot projectively ascribe their properties to objects we could not perceive, such as fundamental particles, which must also have a qualitative nature if Heil is right; meanwhile,

the identity of neurological powers with first-person experience, ostensibly an example of a qualitative object with which we are familiar, is a matter of stipulation on Heil's part. For Loar and Papineau, the use-mention distinction is a way of explaining why physics is silent on the existence of Heil's qualities, and why it does not need them in order to accurately represent objective reality. The very existence of qualities therefore seems to be 'merely' subjective, as to mention an experience makes no essential reference to qualia.

But now it looks as if pointing to the difference between subjective and objective is either trivial wordplay, which is inadequate to dispel the mystery of consciousness, or else a non-trivial but unbelievable, either paradoxical or eliminativist, metaphysics of mind, in which qualities are, objectively, non-qualitative. What are we to make of this?

#### **[4.5.1] Unbelievable Physicalism**

The use-mention distinction is an alternative to Kripke's essentialism, his "semantic premise" according to Loar (600), where two independent concepts can pick out the same property only so long as they refer to contingent modes of presentation. Denying this premise opens up the possibilities of strong necessities. Papineau likewise denies that Kripke's argument against the identity theory has a defensible metaphysical conclusion, only a "psychological" one: "p seems false, so p is false. This is not a particularly strong form of argument. Many true things seem false" (Papineau 2007a, p486). Kripke's argument fails, he thinks, because it is possible for there to be a split between theoretical and intuitive kinds of belief (*ibid* 483). He likens the mystery of consciousness to optical illusions which still fool us after we come to recognise why they do so. Insofar as physicalists cannot explain away their own disbelief in physicalism, Kripke's challenge stands, and this is how Papineau proposes to read *Naming and Necessity's* anti-physicalism, as an *ad hominem* argument to the effect that physicalists

do not really believe their own claims; “Kripke’s argument isn’t that descriptive content is needed for a posteriori identities per se, but rather that it is needed specifically for identities that continue to seem possibly false after they are believed” (*ibid*, p482).

Papineau concedes that physicalists still have work to do in accounting for what else could be going on when we ponder physicalism’s seemingly impossible truth. But he accepts that the answer (whatever it may be) won’t *seem* right, and so abandons the attempt to account for the problematic nature of phenomenal appearances in terms that would satisfy proponents of the hard problem. This smells of irrationalism, however, and plays into the hands of modal rationalists who accuse modal dualists of being intellectually irresponsible. But another version of the use-mention distinction shows how the charge of irrationalism, at least with respect to modality, can be avoided: Christopher Hill (1997) identifies the key premises of arguments for the non-identity of mental and brain states as

Where X and Y are any two properties, if it seems to be the case that X and Y are *separable*, in the sense that it seems to be possible for there to be instances of X that are not instances of Y, then, unless this appearance of separability can be explained away, it really is the case that X and Y are separable... [and] not the case that X and Y are identical. (Hill 1997, p62)

In Boulter’s narrative, this presupposition, which he termed the principle of separability, gained ascendancy via philosophers’ overemphasis on logical possibility. Heil issued a straightforward rejoinder to the (PS) by pointing out that dispositional and qualitative properties do not *seem* separable in our encounters with ordinary objects, even though they can be logically distinguished. By contrast, Hill’s solution denies that we directly perceive the physical at all, drawing on the distinction between sensory, symbolic and sympathetic imagination. Hill argues that although morphological features of brains are directly visible,



and “certain aspects of the structure of individual brain cells” are visible under a microscope, “neither of these things count as perceiving electrochemical activity in living neurons”, being “mediated by theories” (*ibid*, p68).<sup>43</sup> (Since neither symbolic nor theoretical imaginations think their objects under a mode of presentation, and since that is the crucial point here, I will treat them as synonymous from now on.) This is a weaker version of physical concepts, which are not associated by Hill with the way empirical things appear.<sup>44</sup> On Hill’s version of PCS, there is no problematic ‘dual revelation’ of properties because physical properties are not immediately revealed to us. Only phenomenal concepts involve a revelation; the “justification conditions” for self-ascriptions of pain include instances of pain itself (74). That seems to leave us back with something like Lockwood’s view, but if introspection lacks the kind of access needed to reliably discern all the properties of pain, then the certainty of being in pain need not have metaphysical consequences, which is Hill’s position on the issue.

Like Papineau, Hill distinguishes between being in a state and observing it from outside (81); the former leads to modal intuitions about the separability of (e.g.) heat and electromagnetic radiation, whose identity has been “independently established” by the latter. If we follow Chalmers and Goff in lumping sympathetic, perceptual and theoretical imagination together as legitimate sources of insight into modal facts, then we seem able to infer substantive metaphysical truths from conceivability arguments. Since theoretical kinds are conceivably separable from common-sense kinds, this can motivate scepticism about a posteriori identity claims, as knowledge of the scientific facts might only get you as far as thinking the identity of heat and radiation is *contingently* true, leading to scientific anti-

---

<sup>43</sup> Papineau also emphasises the importance of non-presentational modes of imagination: “within the category of imagination, perceptual imagination is not the only kind of third-person imagination: if we can form nonperceptual beliefs and other propositional attitudes about brain states, as we surely can, then presumably we can imagine them non-perceptually too. (Nagel does mention 'symbolic imagination', but only to exclude it from his analysis.)” (1993, p117)

<sup>44</sup> Although Loar himself denies the possibility of a “wider objectivity” (Loar 611) which Nagel hoped could somehow bridge 1<sup>st</sup> and 3<sup>rd</sup> person ascriptions of consciousness.

realism. But the separation of faculties can just as well yield the opposite conclusion, that our sympathetic and perceptual imagination is modally blind outside of limited (everyday) contexts and should defer to science whenever possible, as the success of science has shown how much theoretical imagination can be extended by comparison.

Rather than seeking to explain intuitions of separability by appeal to surrogate possibilities, which would mean identifying scientific rationality with Chalmers' modal rationalism, Hill takes the success of theoretical faculties in the sciences as providing inductive grounds for judging *all* intuitions of separability formed independently of "the relevant empirical facts" (82) as unreliable, regardless of whether these intuitions are formed on the basis of contingent modes of presentation or not (see n.10, p83-4 which describes Kripke's approach to modality as "misguided"). Hill makes no strong claims about what distinguishes empirical from a priori theoretical knowledge, or why a posteriori modal facts are more reliably discoverable via these faculties as opposed to sympathetic imagination. Like Lockwood, he leaves this issue for "another occasion" (Hill n.16 p85). In other words, he does not replace the modal rationalist's deductive-nomological picture of science as needing ideally conceivable logical supervenience with an alternative. Luckily, Wimsatt helps fill in the blanks left by Hill and Papineau about the nature of science, theory and common sense.<sup>45</sup>

#### **[4.5.2] Using versus Mentioning Heuristics**

William Wimsatt goes into more detail than PCS in defining what is meant by 'use' in the 'use-mention' distinction. He defines the 1<sup>st</sup> person point of view as a special kind of heuristic used for interfacing with ourselves, drawing a connection between the concept of a

---

<sup>45</sup> In chapter 6 I will argue that science is special thanks to the importance of symbolic imagination used in formulating scientific theories, and not because empirical evidence itself demonstrates strong identities to us, as Heil claimed.

perspective and that of a paradigm or operation used for approaching certain kinds of problems, e.g. “paradigmatic anatomical, physiological, and genetic problems” (2007, p227) in the life sciences.<sup>46</sup> Perspectives ‘solve’ problems as follows:

Perspectives involve a set of variables that are used to characterize systems or to partition objects into parts, which together give a systematic account of a domain of phenomena, and are peculiarly salient to an observer or class of observers because of the characteristic ways in which those observers interact causally with the system or systems in question. (*ibid*, p227)

Furthermore, “There are things one can accomplish wholly within the subjective perspective, and things that only can be plausibly solved from within the subjective-or a particular subjective-perspective” (*ibid*, p228). For instance, compositionally or interactionally simple objects of enquiry are solveable by one perspective. Perspectives have nothing inherently to do with consciousness; they are just paradigmatic coping strategies attuned to certain problems. But “The smaller kind of perspectives are those things that *look* most subjective, since they are most explicitly keyed to the point of view of a particular kind of organism or observer” (*ibid*, p230, emphasis added). Wimsatt compares them to ecological niches, such as bat sonar. A niche’s relevance to consciousness comes into play when “characterized explicitly cognitively and subjectively, with respect to the cognitive and sensory capacities for and from the point of view of an animal” (*ibid*).

As a kind of “specialist’s knowledge”, the first person is “nearly sealed” (229), leading to the conceivability of idealism or solipsism, and the relative independence of the social sciences from the natural sciences. This explains the autonomy of phenomenal concepts

---

<sup>46</sup> This relates back to his discussion of boundary problems over what constitutes an object; see his contrast between individuating a lump of granite and disambiguating social and psychological phenomena in [3.6], above.

from physical ones. But it also means they can be overextended. Wimsatt points out that some heuristic errors can “affect everything in sight”; it’s “hard to find a place to stand to view it-or to see where it is not. Here a specialist's knowledge may not help, because it may lead us to focus too closely. It is just this kind of mistake we make with our models of rationality.” (24) Since organisms are the antithesis of simplicity, we know, according to this definition of perspectives, that the first person perspective alone can’t capture their essence. As Wimsatt argues,

sometimes problems appear to be big enough, or generally enough stated (e.g., the mind-body problem), that they seem to be intrinsically multi-perspectival. Since a perspective maintains its identity in part by having problems that its corresponding discipline can characteristically solve by itself, the characteristic identification of important problems with certain perspectives and the identity of perspectives tend to break down simultaneously.

(2007, 238)

So we should expect the mind-body problem to motivate overextended heuristics and to raise questions about the integrity or possibility of self-consciousness, and the foundations of philosophy. But such metaphilosophical problems are not solveable without a way of thinking about our perspectives from the outside. The relative independence of the social environment from the material world can be understood, according to Wimsatt, by treating understanding as a tool, which can explain naturalistically the phenomenal/physical concept dualism postulated by PCS.

Wimsatt resists the temptation to arbitrate between reductionism and holism, both of which by themselves would define what things are composed of or grounded in. Both perspectives serve the interests of science, but there is no scientific evidence for whether

parts are more fundamental than the whole (231-5). Indeed, as discussed in chapter 3, Wimsatt argues there are limits to our understanding of how things are composed, beyond which lie the incalculable complexities he terms “causal thickets”. He suggests that “one of the main temptations for vitalistic and (more recently) anti-reductionist thinking in biology and psychology is due to this well-documented failure of functional systems to correspond to well-delineated and spatially compact physical systems” (190). However, organisms can still be given a rough spatial location somewhere between the cosmic and the microscopic. Consequently, “we can understand *in materialistic terms* why compositional relations are problematic” (236). Being unable to say what something is made of doesn’t mean we don’t know roughly how to find out; we don’t need to go ghost-hunting. Subjects must be visible to others as objects, and so presuppose the possibility of being objectifiable under some description, preferably several, even if this has yet to be achieved. “[S]patial objectification is an *active* hypothesis that we apply to those groups of phenomena that tie up into sufficiently neat packages in the right ways ... the mental realm is not denied spatiality, it just has not yet been added to the list.” (192). This is because being possible to situate in space and time is a robust criterion for being real:

it is part of our concept of an object that objects have a multiplicity of properties, which generally require different kinds of tests or procedures for their determination or measurement. It follows that *our concept of an object is a concept of something that is knowable robustly.*” (Wimsatt 197)

Wimsatt argues that “it makes no sense to speak of something as subjective without the other category - which ... involves at least the recognition that there is something outside of the boundary of the subjective” (227). The alternative – that there is no conception of

subjectivity that does not itself use mental concepts – Wimsatt regards as “a bridge that leads to nowhere.” (54) With this he decisively nails his flag to the naturalist mast, as it entails that subjective or phenomenal concepts are necessarily impoverished and unreliable characterisations of their objects. “The terms *first person knowledge* and *third person knowledge* must first be reanalyzed in terms of asymmetries among and limitations on the locational information given by the various sensory modalities” (n.22, p379-80). The personal must be depersonalised, which is to say, spatialized and situated in multiple (second and third-person) contexts that do not deploy its idioms, before picking out a real phenomenon.

Wimsatt’s approach lessens the worry that explaining away subjective qualities as objectively non-qualitative is logically incoherent. What that line amounts to, he suggests, is a prediction that gaining any further philosophical traction on the mind-body problem than we currently possess would of necessity spatialize, and hence naturalise, our concept of mind, because that is what robust grasp of a phenomena has always amounted to in the past: to know one’s way around a subject matter is always to demystify it. Qualia, indeed the concept of phenomenal concepts themselves, would be changed by being recontextualised in such an as-yet-unknown conceptual scheme. But this means that Wimsatt is at least illusionist, if not outright eliminativist, about the mental as we currently understand it, because our current understanding would amount to no more than an impoverished, misleading mode of access to facts about which future neuroscience would have the last word. And so his account ultimately rests on the plausibility of the kind of paradigm shift promised in Churchland (1981). The next chapter therefore seeks, first, to bring PCS in closer proximity with the kind of counterintuitive and revisionist metaphysics of mind that it was originally formulated to avoid; and then to argue, with reference to Raffman qualia, for the conceivability of illusionism about phenomenal consciousness.







## Chapter 5:

### The Disconsolations of Physicalism

#### [5.1] Is PCS Illusionist?

Chapter 2 briefly touched upon the thought that Wilkes's philosophy of mind may be illusionist, and that the same might be said for forms of panpsychism which distinguish 'experience' (qualia that are felt rather than reflected upon, or of which we are somehow unaware) from 'consciousness' proper. Frankish distinguishes two types of illusionism:

Weak illusionism holds that these properties are, in some sense, genuinely *qualitative*: there really are phenomenal properties, though it is an illusion to think they are ineffable, intrinsic, and so on. Strong illusionism, by contrast, denies that the properties to which introspection is sensitive are qualitative: it is an illusion to think there are phenomenal properties at all."

(Frankish 2016, p15)

PCS is best identified as a 'weak illusionist' position: it says that conceptual isolation *misleads* us into thinking phenomenal qualities are non-grainy, intrinsic, ineffable, private and incorrigible, without any of these properties accurately characterising the referent. (In any case, Raffman thinks her argument against Loar's version of PCS undermines the 'strong

illusionist' position that there are only judgements about consciousness, as well; I discuss Raffman again below).

Chalmers has argued that dual revelation, if coherent, will be incompatible with physicalism (we can imagine zombie Chalmers being unable to experience the revelation of phenomenal properties), while less full-blooded accounts of phenomenal concepts don't do justice to the hard problem (we cannot imagine zombie Chalmers lacking phenomenal concepts, hence we haven't accounted for the explanatory gap). Either PCS explains phenomenal concepts and therefore qualia are not *conceptual*, or it fails to explain *phenomenal* concepts because we couldn't have them if qualia were physical. This, in summary, is his "master argument" against PCS.<sup>1</sup>

Carruthers and Veillet (2007, p222) accuse Chalmers of begging the question of what possession of phenomenal concepts amounts to, in favour of sympathetic or sensory imagination: "to introduce the feel of the state into our description of the mode of presentation of Chalmers' concepts is to switch illegitimately to a first-person characterization of those concepts" (*ibid*, p227).<sup>2</sup> They exploit his externalism about the truth of phenomenal judgements, which they compare with a phenomenal zombie's "schmenomenal" judgements. The difference between Chalmers and Zombie Chalmers "is that the *objects* of their knowledge are very different" (*ibid*, p224); phenomenal properties in one case, schmenomenal properties in the other. But the difference is neither here nor there as far as physicalism is concerned, since we can talk about objects without being directly acquainted with them. So physicalism can explain our "epistemic situation" (*ibid*, p221) with respect to the way we theoretically conceptualise phenomenal properties. The

---

<sup>1</sup> Chalmers (2010a) goes into considerably more detail; see especially p312-320.

<sup>2</sup> See also Balog, (2012), p19, who argues that "Invoking the inconceivability of a physicalist account of phenomenal concepts through a reliance on the correctness of the anti-physicalist principles—*the very principles PCS is rebutting*—is far from being a refutation of the PCS. It is a mere refusal to meet the argument on its own ground."

use-mention distinction wasn't supposed to downplay the existence of phenomenal feels, but to explain third-personally the epistemic situation we are in when we conceptualise them. In reply to Chalmers, then, one could pose the equal and opposite dilemma: either qualia are conceptual and therefore reducible to structure, or they are non-conceptual and we cannot be sure phenomenal judgements conceive them correctly (as intrinsic, ineffable, etc).<sup>3</sup>

However, Frankish argues that there is indeed a bait-and-switch in any physicalist approach to consciousness, just as Chalmers suspects. He distinguishes two common conceptions of qualia which philosophers often run together. On the one hand, there is the "classic" variety, which identifies qualia as "Introspectable qualitative properties of experience that are intrinsic, ineffable, and subjective" (Frankish 2012, p668). On the other hand there are "diet" qualia, which are merely "The phenomenal characters (subjective feels, what-it-is-likenesses, etc.) of experience" (*ibid*). It seems clear that if qualia are 'mentioned' at all by third-person ascriptions of phenomenal concepts, then it is diet qualia that PCS is explaining, and not classic qualia. This may seem like a strength, not a weakness, because

Diet qualia look like the theory-neutral explanandum we want. Non-physicalists can go on to argue that the phenomenal character of an experience is (at least partly) determined by classic qualia, which can vary independently of the experience's physical properties. Physicalists, on the other hand, can identify phenomenal characters with representational or functional properties and explain away our intuitions about their intrinsicity – a popular strategy being to appeal to features of our phenomenal concepts. (*ibid*)

---

<sup>3</sup> [3.4.1] used a variation of this argument against Rosenberg.

For instance, PCS is a topic-neutral explanation of the explanatory gap. It has no commitment to physicalism, but no commitment to properties that could resist physicalism either.

At best, however, Frankish argues that this approach succeeds in explaining “zero qualia”: “The properties of experiences that dispose us to judge that experiences have introspectable qualitative properties that are intrinsic, ineffable, and subjective” (Frankish 2012, p669). After all, on the one hand, “Intrinsic properties cannot be explained in functional or representational terms” (*ibid*, p668), whereas “There is no phenomenal residue left when qualia are stripped of their intrinsicality, ineffability, and subjectivity” (*ibid*, p667). In the final analysis, “dual qualia are not explained at all but simply identified with some other feature, and what actually get explained are zero qualia” (*ibid*, p674). Frankish calls this “the dual/zero shuffle” (*ibid*). He concludes that physicalism is better off embracing illusionism about classic qualia and (what amounts to the same thing) defending zero qualia instead.

While promising to save philosophy of mind from “weird positions” (Loar 609), Loar several times raises the issue of empirical evidence inspiring scepticism about qualia; “brain research might discover that (what we take to be) our phenomenal concepts do not in fact discriminate unified physical-functional properties” (Loar 602). He raises the related worry (606) that reference to other minds will pick out brain states that are slightly different to our own. (As discussed in [2.2.2], Papineau is much less optimistic than Loar about meaningful, even if rough, physical similarities uniting mental states across the population; I say more about Papineau’s position below)

Loar responds to his own objection that “if “that type” picks out a physical relation, then the question answers itself” (606), on the basis that “one cannot coherently wonder whether another person in a P state has a state with this phenomenal quality if one

acknowledges that one's concept 'this quality' refers to the property the concept discriminates in oneself (what else?) and that moreover it discriminates P" (*ibid*). He goes on to point out (609) that there is a sense in which wondering about the mind-brain contingency is senseless, if we accept that it is just a fact about e.g. c-fibres that it is like *that* to have them firing. The concept 'pain' just picks out what it picks out, and so cannot fail to refer. Papineau (1993, p137) traces this back to Wittgenstein's aphorism that "You surely know what "It's 5 o'clock here" means; so you also know what "It's 5 o'clock on the sun" means. It means simply that it is just the same time there as it is here when it is 5 o'clock" (Wittgenstein 1953, §350).

On the face of it, this reply seems like cheating. The question was: what are the criteria for being in state P rather than some other state, given that the concepts involved are subjective (*physical* modes of presentation picked out by physical concepts won't help us solve the problem)?<sup>4</sup> More importantly, Loar's reply that such doubts are senseless, because the concept "this" cannot fail to refer, clashes with the direction in which he goes on to develop his position. He writes that "Phenomenal concepts are subjective because they are essentially self-directed" (Loar 609), but does not wish to rule out phenomenal states (which he wishes to identify with physical properties) existing in the *absence* of self-directed concepts. This is because he does not wish to identify the presence of consciousness with the capacity for introspection:

The following could appear possible on my account: another person is in the state that in me amounts to feeling such and such but sincerely denies feeling anything relevant. It apparently has been left open that others have phenomenal states that are not introspectable at will,

---

<sup>4</sup> This question goes for Heil's ontology as well, as he took much the same line as Loar towards phenomenal revelation despite claiming to be a realist about qualia. The criticisms of Heil in chapter 3 support Frankish's contention that there is no middle way between classic and zero qualia. For panpsychists such as Lockwood and Heil who make qualities mind-independent, the slide from classic to zero qualia happens because the qualities end up so mind-independent that subjects could no longer have access to them.

for no requirement of transparency has been mentioned. Then the property that is the referent of my concept of feeling like that could, even if it occurs transparently in me, occur non-transparently in you. But (the objection continues) denying transparency is tantamount to allowing unconscious experiences; and it would not be unreasonable to say that the topic of phenomenal states is the topic of certain conscious states... But it is not obvious that phenomenal properties must be transparent in such a reflexive cognitive sense. What about infants and bats? There has always been a philosophical puzzle about how subtracting reflexive cognitive awareness from phenomenal or conscious states leaves something that is still phenomenal or conscious. But that puzzle is independent of the present account... (611-12)

Loar (605) proposes that identification of mental states in others is like identifying an object of reference in less-than-ideal conditions (e.g. at a distance, at night) when one has previously learned to identify the same object of reference in better conditions. By allowing the possibility of “unconscious experiences” that are opaque to reflexive cognition (a possibility also countenanced by panexperientialists, as discussed in chapter 2), Loar is led to entertain that the self-scanning mechanism by which we form phenomenal concepts in our own case is *not* immune to error through misidentification, so sceptical doubts cannot be senseless after all. He refuses to rule out that “no system of physical-functional properties corresponds to the system of our phenomenal concepts” (612). Worse, the “very possibility ought to make us dubious about the incorrigibility of the judgment that there are real phenomenal repeatables” (*ibid*). Recognising that memory of repeatedly discerning the same phenomenal properties seems to provide strong epistemic warrant for realism about those properties, Loar nevertheless concludes that

if no system of physical-functional properties corresponded to one's putative phenomenal discriminations, an alternative to nonphysical qualities would be this: memory radically deceives us into thinking we discriminate internal features and nonrandomly classify our own states. Strong evidence that no suitable physical-functional properties exist might amaze and stagger one. It would then have emerged that we are subject to a powerful illusion, a cognitive rather than a phenomenal illusion; we would be judging falsely that we thereby discriminate real properties... *The whole point about the phenomenal is how it appears. And that means there is no introspective guarantee of anything beyond mere appearance, even of discriminations of genuine repeatables. (ibid, emphasis added).*

Loar's characterisation of qualia has all the hallmarks of an explanation for zero qualia. Only "positing nonphysical properties" could "forestall the possibility of radical error" – and the supposed advantage of PCS is that it does not resort to positing such properties. Loar's concession to the possibility of radical error about the type-individuation of mental properties demonstrates that, whether we are subject to such a grand illusion or not, he has abandoned the most problematic, and some might say the most obvious, aspects of our access to mental properties (privacy, immediacy, incorrigibility) as illusions.

Katalin Balog highlights these worries about Loar's account of PCS as a motive for adopting her own brand of the strategy, in which phenomenal opacity is ruled out a priori:

Loar thinks of phenomenal concepts as in some way "tracking" their referents. This suggests that he is thinking of the phenomenal concept and its referent as distinct entities related by causation. But it seems that this leaves too much of a distance between, e.g., a phenomenal concept *P* one applies to a particular pain *p* as it occurs and *p* itself, as on this view their occurrence is independent. On a "tracking" view, *P*, or rather, a concept just like *P*, could be tokened by someone in the complete absence of pain... The trouble with Loar's account is

that it opens up the possibility of an appearance/reality distinction for direct phenomenal judgment whereas for direct phenomenal judgment there is no such distinction. (Balog 2011a, p25)

In spite of trying to distinguish the exercise of phenomenal concepts from blindsight, Loar's 'tracking' account does just that: it likens introspection to a kind of blind pointing whose referent can be more authoritatively known from the third person. Balog argues that this characterisation is common among PC strategists, who will claim that phenomenal concepts either lack reference-fixing descriptions or even modes of presentation (2011b, p304). Instead, phenomenal concepts refer via their causal connection with the referent. In other words, brain functions *occasion* the application of PCs. This approach is noncommittal as to how much access we have to what inspires us to use phenomenal concepts. Their denotation is just whatever physical function causes us to apply them in the first place.

Balog contrasts this position to her own 'constitutional' brand of PCS where the referent is directly present in the concept, and is the cause of its own application (self-causation is not the same as a reference-fixing causal chain). "If this account is right, phenomenal concepts have very special *realization properties*: the neural states realizing these concepts are the very same neural states the concepts refer to" (Balog 2011a, p25). This has shades of Russellian monism's appeal to qualia as realizers of physical properties. "Metaphorically speaking, a token of the reference provides the ink in which the token concept is written." (Balog 2011b, p306). Balog's account is still "silent about the nature of phenomenal states" (*ibid*), however. She admits that "no actual account has been proposed of how a concept can be like that" (2011a, p32), and elsewhere briefly mentions (but does not address) Joseph Levine's complaint that "it is impossible to explain cognitive presence by physical presence" (Balog 2012, p7). How could a concept *incorporate* a property? Frankish



agrees: he points out that “Incorporating iron filings into [a brain state] wouldn't make it represent iron” (Frankish 2016, p35-6).

Besides these objections to PCS from fellow physicalists, Goff has complained that constitutive PCS trivialises phenomenal insight into the nature of experience— it fails to give us “rational certainty”, whereby a subject can “rule out any scenario where p fails to be the case” (Goff 2015b, p130-1); “It is one thing to think that a judgement cannot fail to be true; it is another thing to think I can be *justified* in not entertaining any doubt about the judgement” (*ibid*) (He allows that Balog’s brand of PCS can account for “phenomenal certainty”, the certainty that one is *having* a certain kind of experience). Balog agrees that “If phenomenal properties are, as the physicalist claims, physical or functional properties, then there is a clear sense in which acquaintance *doesn't* reveal their nature” (Balog 2012, p15). Echoing Loar’s claim that the “mere appearance” of phenomenal properties gives no deep insight into their nature, Balog writes that “it would be better for the physicalist to analyze acquaintance and the substantiality of phenomenal belief in terms of the phenomenal presence of the introspected properties in phenomenal judgments; and not in terms of our direct grasp of the essence of phenomenal properties” (*ibid*). But this makes it sound like qualia are an artefact of judgement – that intrinsic, ineffable, and subjective properties only *seem* to be present. Being *experienced* as having these properties could just be a contingent manner of conceiving the physical, rather than indicating the direct grasp of the property itself.<sup>5</sup>

In their introduction to *The Nature of Consciousness*, the editors Block, Flanagan and Güzeldere call the “image” that accompanies phenomenal concepts a “mode of individuation”, not a mode of presentation (p.xx). Confusingly, an image can mean a

---

<sup>5</sup> Echoing Lockwood’s subject-deflationism, Balog (2011a) starts with a quote from Imre Kertész to the effect that the eye [PCs] can’t see itself seeing [what it’s made of], imagery which is suggestive of illusionism.

“secondary version” which *triggers* a concept, or which “quotes” it, rather than directly presenting its content. Papineau’s “versions”, which take the latter approach, are just the zero qualia version of constitutive PCS. But what, in the end, distinguishes this from Balog’s supposedly more accommodating line? She cannot reconcile what qualia-realism demands with the demands of physicalism. As Raffman puts it, the supposed advantage of PCS is that “Nothing is red — objectively, phenomenally, or otherwise. The brain isn’t red, and the introspective concepts neither are, nor need otherwise introduce, phenomenal red. Just as tokens of the English predicate ‘red’ can represent objects as red without themselves being, or otherwise needing to present, red” (Raffman 1995, p298).

However, as section [4.3.4] discussed, Raffman thinks she has an empirical case against PCS. As I concluded there, Raffman is not exactly favourable to anti-physicalism in spite of this. But it is not clear that she has a knock-down argument against PCS either. Her approach can just as well be seen as a modified version of the doctrine, albeit one which underscores PCS’s otherwise implicit illusionism.

## **[5.2] Raffman and Balog on “Presentational” Concepts**

The difference between Papineau’s and Balog’s versions of PCS seems, on its face, to be a difference between Papineau’s quotational concepts, which are non-presentational in that they avoid literally locating phenomenal properties in the brain - - and an account in which phenomenology itself is really constitutive of PCs (a presentational account). If Frankish is right then a presentational concepts strategy is impossible, because it presupposes classic qualia and so cannot be physicalist. Raffman claims to take the opposite line, arguing that empirical evidence supports a presentational account of qualia, *and* that this account is both

consistent with physicalism and refutes PCS. I will suggest that the dispute is largely semantic. But apart from the travails of the concept of 'presentation' in this context, there remain questions as to what PCS itself amounts to. Balog (2011b, p296) writes that "If qualia represent then it is plausible that they represent non-conceptually. That is, they do not have language-like structure but rather are akin to pictures and represent in something along the lines of the way pictures, images, graphs, and other so-called analogue representations represent." One wonders what has become of the phenomenal *concept* strategy, here. She proposes an alternative, whereby qualia may *not* be "intrinsically representational. A way of thinking about this is to say that they are aspects of sensation and can occur in the absence of representational content" (*ibid*). But again, if true, there is the issue of what is captured by phenomenal concepts: are we to suppose it is the non-representational (but conceptual?) sensation itself?

A plausible candidate for Balog's characterisation of phenomenal concepts/properties is Raffman's notion of a "presentational representation": "A presentational representation is one that does something more like showing than telling, more like exemplifying than merely standing for, more like instantiating than merely designating, its content. Here is a presentational representation of (let us suppose) red<sub>31</sub>: •" (1995, p298). Raffman identifies the possibility of presentational representations with the "traditional qualia view" which believes in "intrinsic redness" (*ibid*). By contrast, she says, PCS "rejects any presentation of phenomenal colour... the perception is just a 'physically acceptable' state of the brain to which first-person concepts are applied in introspection" (*ibid*). It is the presentational account that Raffman favours: "We have phenomenal *representations* of Red<sub>31</sub> — namely our perceptions of Red<sub>31</sub> — and *nonphenomenal* concepts of Red<sub>31</sub> and Red<sub>31</sub>-perceptions; but no phenomenal concepts of these fine-grained values" (*ibid*, p299). (Raffman's terminology is confusing because phenomenal

representations are really presentations; calling them “presentational representations” simply adds confusion) She also claims not to “take sides” (*ibid*, p294) with respect to the “existence and nature of qualia”, but immediately qualifies her position of neutrality with the sceptical speculation that “there may be no subjective facts”. So, ‘presentation’ in this context need not imply intrinsic properties in the sense of Frankish’s “classic” qualia after all; as we will see, Raffman’s position is in fact thoroughly illusionist.

Meanwhile, Balog is strikingly sanguine about the prospects of PCS being able to accommodate Raffman’s criticisms when she refers to the fineness of phenomenal grain in an overview of PCS as among the ‘desiderata’ it must explain or “explain away” (desiderata which receive no further comment in Balog’s paper):

*(f) Fineness of grain. There is a fineness of grain in experience that cannot be captured by the phenomenal concepts possessed by the subject of experience; certainly not by concepts that can be applied and reapplied in thought. We can discriminate between millions of different shades of colour experiences, but we can only form at most a few dozen standing colour-experience concepts.*

*(g) Semantic stability. PCs refer to the same properties independently of the actual context; i.e. their extension can be determined independently of any empirical discoveries. In contrast, the reference of semantically unstable concepts (like e.g. water, which refers to the liquid, transparent, etc. stuff that the thinker is in contact with) is actual-context dependent.*

(2011b, p299)

According to Raffman, (f) and (g) are deeply in tension with one another; fineness of grain entails semantic instability. Block, Flanagan and Güzeldere (1997) characterise PCS as accommodating the thought that Mary the colour scientist learns new facts upon seeing

colours for the first time, just as anti-physicalists believe, but only about old properties which she has merely learned to conceive in new ways (p.xx). However, the semantic instability of Raffman qualia render it unclear what exactly Mary has learned about the physical properties of her brain by using, as opposed to mentioning, colour concepts for the first time. The fact that red is like *this* is hardly informative if 'this' can, without one's first-person awareness, be predicated to any number of finely-distinguished thises. Which is why Raffman characterises Loar's view as predicative rather than demonstrative, in spite of his claims to the contrary.<sup>6</sup> He makes it sound as if qualia are exhausted by recognitional concepts or mentalese predicates.

Raffman's attack on phenomenal concepts includes the idea of conceptualising phenomenal consciousness in 'mentalese' as much as in natural language: "the problem is not that you can't report the determinate content of your introspection because that content is not expressible in natural language. Rather, it's that you can't report that content because you can't recognize it when you have it; you can't learn to say so much as, 'There it is again'" (1995, p300). But by this Raffman seems to have in mind mentalese along the lines of Papineau's 'secondary versions'; mental mechanisms that somehow 'quote' qualia without presenting them. Since, according to Raffman's confusing lexicon, a representation can be 'presentational' of its content, Balog's accommodation of non-conceptual experience would seem to reconcile PCS with fine-grained qualia after all. If Raffman thinks we have phenomenal representations but not concepts of Raffman qualia, then perhaps PCS should

---

<sup>6</sup> Raffman confusingly introduces demonstrative forms of PCS later in the paper (1995, p302-3), which she associates with Terence Horgan (1984), not Loar. She immediately dismisses it for the same reason as non-physicalists tend to: "the demonstrative line has it that we introspectively think of all of our experiences as, simply, *this* (property or experience). In other words, it renders all experiences introspectively identical" (1995, p303). The only way out, she argues, would be to reintroduce descriptive content pertaining to relations of resemblance among hue, shape, pitch and so on in order to "beef up" the representation; this was Schroer's response to 'thin' demonstratives discussed in the previous chapter. According to Raffman this entails that insofar as the descriptive account is coherent at all (that is, insofar as it avoids the differentiation problem), it is really just the predicative kind. I return to Schroer's position below.

be rebranded the phenomenal representation strategy. Why, then, is Raffman so sure that all forms of PCS are ruled out?

Raffman argue that “The cornerstone of [PCS] is a distinction among different ways of knowing or representing physical facts. One of these ways of knowing, namely the first-person introspective way, is supposed to capture how I myself represent my own first-order perceptual states” (1995, p300). Raffman denies that there are two ‘ways of knowing’ at all, thus ruling out PCS’s dualism of non-overlapping physical and phenomenal concepts. But (at the risk of repetition) it is difficult to see how the sheer possibility of representing first-order perceptual states to oneself does not qualify as a ‘way of knowing’ in some sense which a suitably enlightened PCS could not accommodate. However, it turns out Raffman also (in some sense) denies *first-order* perceptual states are represented at all, let alone to oneself; she could plausibly qualify as an eliminativist about the self, for all the relevance selfhood has to fine-grained qualia. As if this were not confusing enough, Raffman clarifies her radical position in the course of reconsidering a variant of demonstrative phenomenal concepts, which finally receives a fair hearing after previously downplaying PCS’s association with demonstratives (see n.6, above). Perhaps, Raffman supposes, the differentiation problem raised by the concept ‘this’ always being identical despite being triggered by all the different kinds of experiences – being too thin to characterise them – is solved because “The neural state is the object, and the demonstrative the ‘mode of presentation’, of introspection.” In other words, “our introspective ways of thinking are differentiated by the differences among their referents: Our introspective states are differentiated by our introspected states” (1995, p303-4). Rather than rejecting the demonstrative line once more, Raffman agrees that differentiating phenomenal contents by their extensions rather than their intensions is promising, with the following caveat:

citing the instantiation of the relevant physical property as referent does fill the informational gap — but only from a third-person point of view, as it were. The trouble is that the physical property instantiation cannot supply the requisite content *from a first-person introspective point of view*. Or, better: although it supplies the requisite content, it does not supply it under the right mode of presentation; it does not supply that content under a mode of presentation that makes it accessible by introspection — in either Mary or her audience. (*ibid*, p305)

Raffman is replying to Horgan (1984) who thinks the information expressed by PCs can only be “conveyed” by having the experience. *Raffman inverts this claim*: she thinks we *don't* grasp the essence of a property from the first person at all. Raffman isn't suggesting that we have some other, non-conceptual mode of access to those states the way we do to real objects — *except* insofar as we consist in sub-personal mechanisms that can *only* be conceptualised in third person terms:

I do not deny that perceptual and other mental contents could be rendered in predicative or other coded form at a 'sub-personal' level of processing. For all I have said, predicative modes of presentation may be the brain's standard currency. (What that would mean, of course, is that sub-personal parts of the brain are making sub-personal type-identifications.) My point is only that predicative modes of presentation cannot be the currency of personal-level introspection: they cannot be the way *you* represent your own mental states *to yourself*. (1995, p300)

This is a very strong form of metaphysical realism about Raffman qualia. So she thinks there is a kind of appearance-reality distinction to subjective experiences; how else could subjective experience be independent of the subject who has it? But Loar already allowed

for this, as it was his basis for extending phenomenal experience to creatures that lack phenomenal concepts. Far from being incompatible with PCS, Raffman's position is simply the apotheosis of the paradox I have argued runs through both panpsychism and physicalism: that the mind is a mind-independent thing.

One way of looking at it, then, is that Raffman qualia actually only refute constitutive PCS, because they aren't constitutive of colour *concepts*. Ironically, Raffman's position seems closer to the 'tracking' form of PCS to which Balog objected. Recall Loar's sceptical worries about the possibility that "no system of physical-functional properties corresponds to the system of our phenomenal concepts". To be sure, Raffman denies that phenomenal judgements *fail* to discriminate; rather, they're just not discriminate enough to account for the *content* of experience itself. She also frames her argument in a different way, as targeting introspectable type-identification: Raffman is talking about *failing* to conceptualise experiences we have, whereas Loar's example is one of conceptualising, erroneously, what we *don't* in fact have. But in the end, is this an important distinction? Just as, on some versions of PCS, our concepts pick out their referents via a causal connection from the latter to the former, so Raffman qualia are opaquely causally connected to introspection. Experience always contains, among other things, more shades of colour than we could recognise; for experience to lack Raffman qualia altogether would be for it to lack colour properties in the first place. Yet the grain of experience is inaccessible to introspection because it is essentially pre-reflective.

Metzinger, in his presentation of Raffman's argument, goes so far as to argue that Raffman qualia are consequently not *phenomenal* properties at all, on account of the fact that they cannot in principle be identified by phenomenal concepts (2003, p72-3). Only third-person physical concepts capture them. To the extent that Raffman qualia are present to us at all, then, their mode of presentation is physical, not phenomenal. Raffman seemed



at first to be suggesting that there is access to first-order states besides demonstrative phenomenal concepts. But seen in light of the above, what she is really saying is that ‘we’ don’t have access to those states at all: parts of our minds are asubjective, and only physicalism can fill out what this means. The thought would seem to be that Raffman qualia are not even first-order states that happen to be inaccessible to second-order judgement: they are *zero* order states, which sounds almost as peculiar as Rosenberg’s level zero (non)entities existing outside of any determinate context.

As usual, Wimsatt helps clarify another seemingly nonsensical piece of naturalist metaphysics. In the course of defining “perspective” – which to repeat, is for him a kind of tool for systematically accessing and intervening in a reasonably well-defined range of physical properties – Wimsatt notes that they can sometimes be separated into “levels” (Wimsatt 228-9). Levels are posited when perspectives “decompose systems in ways that... are hierarchically rationalizable relative to one another, so that the parts of one perspective are all whole systems in another (in which case the perspectives are related to one another as different level descriptions of the same system) (*ibid*). As I noted in [4.5.2], Wimsatt thinks of the first person as a “nearly sealed” level, one whose domain – the objects of introspection – is isolated enough to encourage metaphysical reifications of the mental. The qualification ‘nearly’ comes to the fore when considering Raffman qualia; these are an apt example of example of what Wimsatt terms “level leakage”, where a domain of information which some perspective has no traction on nevertheless has causal effects on that perspective’s operations. Such effects are not felt *indirectly* from outside some sort of well-defined domain, but are rather a direct reminder that no perspective has complete integrity, e.g. “functional systems are still subject to physical, chemical, and biological constraints at a number of levels, and never completely lose the marks of the systems from which they have evolved” (Wimsatt 190). Raffman qualia are *there* amongst the objects sorted by

phenomenal concepts, but are themselves only sorted by physical concepts (as in [4.5.2] we must keep in mind that Wimsatt conceives the first-personal in an essentially impersonal way, a “slice” of the organism).<sup>7</sup>

Nevertheless, and at the risk of repetition, all of this is just a variation on qualia being strongly identical with physical properties, just as PCS claims, only with the caveat that phenomenal concepts misrepresent our epistemic situation. Hence a reconciliation between tracking and constitutive brands of PCS is possible: it could be ‘zero qualia’ that are directly presented, but *misrepresented* (tracked) as classic qualia. Loar, to repeat, may even have had something like this in mind with his “cognitive rather than phenomenal illusion” of discriminating “real properties”. Similarly, while Raffman’s 1995 paper “On the Persistence of Phenomenology” claims to remain neutral on the hard problem of consciousness, the final chapter of her earlier book *Language, Music and Mind* (1993) is much more forthrightly illusionist. Here Raffman advances the hypothesis that the apparent resistance of qualia to being analysed in structural or functional terms (1993, p144) might be due to their “cognitive impenetrability” (*ibid*, p133).<sup>8</sup> This chimes with “the basic illusionist claim that introspection delivers a partial, distorted view of our experiences, misrepresenting complex physical features as simple phenomenal ones” (Frankish 2016, p18). But it is also in keeping with PCS, which seeks to explain away the appearance of mind-brain contingency as a consequence of self-scanning mechanisms causing us to conceptualise experience in a special, conceptually isolated way. Raffman writes that “*qua* perceivers we are so designed that the grain of conscious experience will inevitably be finer than that of our [conceptual] schemas” (1993, p136). Another way of thinking about this is that the very non-

---

<sup>7</sup> Wimsatt (229) speculates that some sort of leakage of the impersonal into awareness would be necessary to explain where we ever got the idea of mind-independent reality and the existence of other minds. The idea that our knowledge of these topics is strictly ‘impossible’ from within the first-person is explored at length in Valberg (2007).

<sup>8</sup> Recall that in chapter 2, the cognitive impenetrability of sensations was argued by Peter King (1994) to have sunk Medieval Scholasticism’s aim of explaining the mind’s access to universals.

conceptuality of consciousness rules out knowing whether or not it is physical; which puts her position closer to Balog again (and Heil in his more physicalist moods). Phenomenal concepts were never meant to provide an analysis of their referents, anyway – only physical concepts could do so, according to PCS.

Raffman would no doubt reply that our being unable to “say so much as, 'There it is again'” (Raffman 1995, p300) about Raffman qualia rules out even the most minimal, uninformative self-*predication* of phenomenal concepts. However, it doesn't follow that “I cannot represent my own experiences to myself, in introspection, as RED<sub>31</sub>-EXPERIENCES.” (*ibid*); she admitted we can represent sub-personally. Raffman seems to have a quite sophisticated form of introspection in mind. What about the thought that qualia are ‘self-illuminating’? She is, after all, saying that we *are* pre-reflectively conscious of Red<sub>31</sub> (“first order states”). She denies sub-personal representations are conceptual (*ibid*, p299-300); but again, she seems to have a sophisticated version of ‘concept’ in mind, when a less committal PCS would be compatible with them. Papineau, for instance, denies the existence of an ‘inner light’ by which phenomenal states are made conscious (see below), and his account of PCS does not presuppose a person ‘for whom’ qualia would be illuminated as ‘THAT’. Raffman characterises Papineau's position as one in which “introspection is ... the application of a demonstrative concept ... to some 'physically acceptable characteristic' of the brain” (*ibid*, p303), and complains that the concept would fail to determinately capture its referent. But this is unfair. Why couldn't Papineau claim the demonstrative is a function of sub-personal representations directed at presentational content, which is what Raffman is committed to? Isn't the point precisely that Raffman qualia are strongly identical with subpersonal processes, just as PCS claims?

Similarly, if adequate descriptions of our perceptual states are only available using third-person concepts, the apparent contingency of first-person experience vis-à-vis the

brain could still be explained as an antipathetic fallacy. The fallacy would just be one of assuming consciousness cannot be represented as an object from the fact that we do not represent it in any particular way to ourselves, rather than due to a dualism between physical and mental concepts. Not knowing *how* we are representing experience just from introspection is precisely what PCS claims: it just says (implausibly) that this is due to the way we conceptualise qualia, rather than the way we *don't*. Critics of PCS point out that its analysis of how we think about experience is inadequate because for PCS to be true would presuppose a revisionary account of qualia, on account of 'this' being too thin. But perhaps it should not be thought of as seeking to characterise qualia anyway. If what we experience is only differentiated insofar as its mode of presentation is physically conceptualisable, then why can't the phenomenal side of our concepts be undifferentiated? Rather than a theory of how we introspect, why couldn't the demonstrative be a placeholder for our *lack* of concepts for what appears to us?<sup>9</sup> Raffman's disagreement with PCS is only pertinent with regards to the form taken by third-person theories of why we think phenomenal experience poses a problem. From the first person we are in the same situation either way – we know we're having an experience, but we can't say *what* experience is. On PCS this knowledge is foreclosed from *without* – since in deploying phenomenal concepts we have no access to physicalist concepts of the same presentations, and vice-versa (there is no binocular picture available to us). For Raffman the foreclosure comes from within – phenomenal experience outruns the concepts we use to represent it to ourselves.

I conclude that Raffman's position is compatible with PCS – but only to the extent that both are more illusionist than they tend to let on. But isn't illusionism about

---

<sup>9</sup> Recall that in chapter 3 I argued that the same could be said in defence of agnosticism on the question of whether reality has a qualitative dimension. That physics is silent about the existence of qualities, as Heil put it, need not be understood as evidence that it only "partially conceives" things in themselves that are better conceived by metaphysicians. Rather its silence could be a matter of modesty given that nobody, and especially not philosophers, has the kind of access to reality needed in order to say more.

consciousness a crazy position to defend, a *reductio ad absurdum* for physicalism? If illusionism is supposed to be compatible with a hard-line naturalist epistemology, in which science is the exclusive arbiter of what there is, then the answer may well be yes. However, another way of looking at it is that illusionist arguments reopen the possibility of panprotopsychism, neutral monism, and constitutive panpsychism. All of these were previously written off by Goff for being obscurantist, being poorly motivated, or (most tellingly) for being little better than physicalism anyway. But the first two objections can be levelled against Goff's speculations as well. That leaves only the question of what separates a sufficiently reductionist neutral monism from physicalism, which has long been a fundamental issue for Goff, to whom I now turn once more.

### **[5.3] Partial Revelation and Selfhood**

Firstly, and to rehearse some of the themes discussed in [4.3.1], let us remember that Goff objected to PCS on the grounds that it cannot account for (1) phenomenal certainty, in which we know for sure *what* phenomenal quality is instantiated; (2) phenomenal insight, in which we know *much more* about phenomenal qualities than simply that they are present to us, and (3) rational certainty about resemblance relations *between* experiences, such as red's proximity to orange. These 3 properties are interconnected, so even if one or more could be accounted for, our access to experience could still not be considered rational. At best, PCS can say we pick out phenomenal qualities (a) "by description", (b) "with radically opaque concepts", or (c) "with some hybrid of descriptive and radically opaque sub-concepts" (Goff 2015b, p133-4). If (a) or (c), PCS cannot account for (1); if (b), PCS cannot account for (2) or (3). A radically opaque concept would be something like 'that', which

would give us no insight into how experiences are distinguished from one another, and consequently make phenomenal certainty vacuous and non-rational. Knowing experience by description would be more like knowing its reputation than knowing the real thing (*ibid*, p131). Meanwhile, a hybrid of opaque (sub-) concepts would make it possible to be wrong about what *exactly* we are experiencing; our qualia could be wiggling slightly without us knowing that any change was occurring.

In response, Goff has argued for the “consciousness constraint”: “Any adequate theory of reality must entail that at least some phenomenal concepts are satisfied” (2017a, p3). This entails transparency for simple experiences under normal conditions:

This principle does not entail that we never make mistakes about consciousness, even in one’s own case. Consider the following example. You are extremely anxious during a visit to the dentist. The dentist applies pressure to the inside of your mouth. The combination of anxiety and pressure leads you to think that you are feeling pain when in fact you’re not. This may be a case in which you mischaracterize your own conscious experience: you think you are feeling pain when in fact you are feeling anxiety and a sensation of pressure. The Consciousness Constraint tells us merely that *some* phenomenal concepts are satisfied and so is consistent with the fact that we apply the wrong phenomenal concepts in certain cases. (*ibid*)

Citing the example of pain, Goff points out that “We are not dealing with complicated mathematics here” (2014b, p12). The refutation of physicalism turns on simple insights in which there is no realistic possibility of confusion.

The devil is in the details, however. Evidence from the neuroscience of pain suggests it may be a particularly bad example of a property about which we suffer no confusion. As Nikola Grahek (2007) observes, pain “is actually a complex experience comprising sensory-

discriminative, emotional-cognitive, and behavioral components” (p2). Paradigm cases of pain, then, would really be a confusion of those components, which can be introspectively discriminated, as demonstrated by the introspective reports of patients with pain asymbolia, in which cognitive and behavioural components have been lost due to a brain lesion. Patients ‘suffering’ this deficit will insist they feel pain as intensely as before, only minus the painfulness: they will no longer find the sensation unpleasant or seek to avoid it. It is difficult to imagine what such an experience could be like, and it seems at least some patients agree: “Schilder and Stengel [1928, p. 151] reported that when their patient was pricked on her left hand and asked whether it hurt, they received the following reply: ‘It hurts indeed, but I do not know what that really is’” (Grahek 2007, p45). It seems that in such rare cases of ‘pure’ pain (*ibid* 36-7), the sensation “comes to such sensory indeterminacy that it cannot be distinguished from other unpleasant sensations, or sensations of other quality, and loses all informational power with regard to the location, intensity, temporal profile, and nature of harmful stimuli” (*ibid*, p2). In other words, here we have a quale to which a thin demonstrative is entirely appropriate, despite pure pain, according to Goff, providing the most vivid example of what PCS fails to account for. This supports and extends in Schroer’s transparent account of consciousness, which he applied to colours; pure pain seems much like Schroer’s hue-in-itself, a property made vivid only by its relational properties with other components.

Goff contends that “A normal individual in agony can know for certain that she is in agony, not merely that some aspect of her agony is instantiated” (Goff 2015b, p135). But insisting that an individual qualify as “normal” in order to rule out physicalism brings us back to the issue of Goff’s fallibilism, which was previously brought to attention in [4.4]. In Goff (2012), he wrote that “if something is fully conceivable, and our best efforts of reflection do

not reveal any incoherence, I think this is very good grounds for concluding that what we are conceiving of is ideally conceivable” (p743). By contrast, PCS

puts introspective beliefs about phenomenal qualities in the same epistemological boat as ordinary perceptual beliefs: reliant on the proper functioning of an information-gathering mechanism. My belief that there is a table in front of me is reliant for its truth on the proper functioning of my senses... [but] the belief that I am currently experiencing phenomenal orange is not subject to doubt in the way that my belief that there is a table in front of me is. (Goff 2015b, p133)

It seems clear that our beliefs about having sensations are *more* reliable than our beliefs about the external world. But Goff is hesitant to put them in another category altogether. Confusingly, he says in the same paper that “our reason to believe in real acquaintance is somewhat similar to our reason to believe in the external world” (*ibid* p130). This is a remarkable concession from a philosopher willing to disregard common sense altogether in favour of conscious pillars of salt. After all, if there is any room for weighing up one’s beliefs against what is reasonable, or what seems evident, then most of Goff’s metaphysics will have to be set aside.

Goff admits several times in his (2017a) (see esp. p230-1) that his position involves an element of ‘noumenalism’, whereby certain crucial details of how subjectivity is produced are said to be beyond our ken. In Goff’s case these details concern the deep nature of spatial relations, and the way in which the cosmopsychic decomposes into individual subjects. On the face of it, noumenalism would seem to support mysterian or illusionist physicalism as readily as it does cosmopsychism, and at much lesser ontological cost. Perhaps this explains why Goff has repeatedly engaged Schroer’s argument for phenomenal translucency, and



why he was so worried about phenomenal bonding. In Goff (2009a), he admits that the bonding problem introduces an element of opacity into subjectivity which seems to support mysterian physicalism as strongly as it does panpsychism. Starting from the observation that we “have epistemic access to only one subject of experience ... via introspection” (*ibid*, p132), and are unable to identify subjects qua subjects via sense perception – as both mysterians and PC strategists as well as anti-physicalists agree – Goff concludes that we have no faculty for conceiving how subjects relate to one another, and thus no solution to the bonding problem. Granting that there is such a solution, however, leads to the following dilemma: the mysterious relation capable of bonding micro-phenomenal properties or subjects into o-subjects could just as well be one relating unconscious physical entities, and thus support emergentist physicalism.

The non-panpsychist theorist who postulates the physical-to-phenomenal bonding relation to explain consciousness must confess to a certain degree of ignorance as to how exactly non-conscious particles sum together to make subjects of experience. But similarly the panpsychist who commits to the phenomenal bonding relation must confess to a certain degree of ignorance as to how exactly little subjects of experience sum together to make human and animal consciousness, which is after all the kind of consciousness we have a pre-theoretical need to explain. It is not obvious that the former kind of ignorance is any greater than the latter. (*ibid*, p134)

It was only after this crisis of faith that Goff began to shift to defending consciousness\* against arguments for partial revelation (see [4.3.3]), and to defend cosmopsychism against reductionism. For if o-subjects are grounded in a greater whole rather than in a mysterious sum of their parts, then nothing about how we are put together would be hidden from

introspection, and perhaps more importantly, nothing about the mysterious wider context in which consciousness is subsumed could recontextualise what we introspect, such that physicalism would suddenly be conceivable again.

Goff objects to translucency, whereby some aspects of consciousness are foreclosed to us, on the grounds that it is a slippery slope back to mysterianism or illusionism. This might seem a little hasty, as it earlier looked as if insight into simple experiences, like red or pain, were enough to fend off physicalism. But irreducible subjects and irreducible experiences may be mutually supportive. Raffman qualia, and the case of pain asymbolia, give us reason to believe there is more to phenomenal experience than the introspecting subject is aware of. By contrast, if the part of experience that I am certain of could be clearly delineated, then the worry that some aspect of it escapes awareness would be neither here nor there. This was Goff's response to Schroer: "we could simply substitute the word 'consciousness' in what follows for 'consciousness\*', defined as 'that aspect of consciousness we understand the nature of a priori'" (Goff 2014a, p80). Knowledge of subjecthood needs to be fully transparent and a priori in order to ground phenomenal revelation of even simple properties like red, or pain. In his book, Goff formulates irreducible subjecthood as an extension of phenomenal transparency, "the thesis that direct phenomenal concepts reveal the essences of the states they denote" (2017a, p178).

I take it that subjecthood is a determinable of which each conscious state is a determinate. For example, to be pained is to be a subject in some specific way; to have an experience of orange is to be a subject in some other way. I further take it that if one grasps the essence of a given determinate one thereby grasps the essence of the determinable of that determinate. For example, I couldn't understand what it is for something to be spherical without grasping what it is for something to be shaped, or what it is for something to be red

without understanding what it is for something to be colored. It follows that, for any phenomenal property I conceive of under a direct phenomenal concept, in grasping the nature of that property I thereby grasp the nature of subjecthood. (*ibid*)

Goff's emphasis on our grasp of subjecthood casts back to his dispute with constitutive panpsychism, which, as chapter 2 revealed, has repeatedly shown a tendency to posit conscious states existing in the absence of subjects; Lockwood's disclosure view, Rosenberg's panexperientialism, Chalmers' panqualityism. And ruling out those philosophers' subject-deflationism makes it much harder to naturalise consciousness. But even if we grant him that point, on the basis that there is something suspicious about mind-independent mental qualities, nevertheless Goff's concept of what a subject is is too reificatory to be realistic. Either it is a composite object comprised of a group of experiences, a picture strongly suggested by Goff's formulations of, and solutions to, the combination problem, or the subject stands as something like a universal in its relations to particulars.

With respect to the conception of subjects as composite objects, Goff is far too quick to conclude that we grasp their nature by virtue of grasping their parts. Introspection does not clearly reveal the extent of the present moment, the nature of time, whether we enjoy personal identity over time, or whether we possess free will. And Goff is willing to defer to science in order to answer those sorts of questions. He is on much shakier ground in thinking that we have complete, or close to infallible, knowledge of all that we are experiencing at any moment, than in thinking our knowledge of boilerplate examples such as spheres or the colour red is complete. But Goff's arguments for irreducible o-properties never established that human subjects consist of one well-defined bonded set of experiences. What they did suggest was that experience cannot be carved up into neat, separate sensations or

properties: experience is not like a cloud of microphenomena. But conversely, it does not seem to be like a kind of giant atom, with fully determinate boundaries, either.

With respect to the second option, of viewing our grasp of subjecthood as the grasp of a universal, this calls to mind the mysticism of Medieval Aristotelians such as Dante, discussed in chapter 3, where the form of subjecthood was common to all beings, beings which were emergent from the subject, rather than subjectivity emerging from material being. And one of the motivations for Goff's cosmopsychism was to avoid postulating seemingly arbitrary boundaries separating living organisms from inanimate material, in order to avoid brute emergentism. There are important differences, however, as the coherence of the Medieval view was dependent on everything being a mode of a universal, the "thought that thinks itself", which was modelled on the rational intellect's ability to overcome the limitations of the senses, a point carried over into Cartesian metaphysics.<sup>10</sup> To the extent that Goff's concept of mind is grounded in sensations, and uses irreducible subjectivity as a crutch to shore up the irreducibility of sensations, the old Aristotelian picture will be of no comfort. If anything, embracing a form of subject-deflationism would be more congenial to Goff's sceptical outlook on personal identity even in light of his cosmopsychism. He could deny that the scent of a rose is further decomposable into simpler qualia, and yet deny that there is anything it is like to be 'me', except in the trivial sense that I am always somebody. This would at least account for the inescapable possibility of vagueness and error in one's phenomenal judgements as to what set of o-experiences comprise oneself at any time.

Surely, however, there is *something* about consciousness that we're certain of, that is not predicted by physicalism or subject to scepticism, and that isn't simply an illusion? While sensations provide an interesting focus for analysis of the mind-body problem thanks to

---

<sup>10</sup> That sensations, once properly understood, would turn out to be more like thoughts than qualities, is also strongly implied by Wimsatt's illusionist approach, though he also emphasises that thought is a matter of achieving robust modes of calibration with the environment rather than a grasp of universals.

their lying at the boundary between the subject and world, and being open to scientific study as much as to introspection, for this reason they cannot help shore up Cartesian certainty: perception is by its very nature translucent, as opposed to transparent. Hence Raffman's conclusion that qualia are asubjective physical properties, making her an illusionist twice over: she denies we have pre-reflective first-personal certainty about qualia, since their phenomenal properties are just cognitive illusions, *and* she denies we have reflective incorrigibility because there is no self which experiences Raffman qualia. Reflection can't be the means by which we discover or even generate certainty with respect to phenomenal properties, because they are not the sorts of things we can *be* reflectively certain about. But perhaps that just means something besides phenomenal properties will have to be the source of our certainty – bearing in mind that they played no significant role in Descartes' discovery of the cogito at the end of doubt.

That something is provided by Keith Gunderson (1970). Wimsatt (379-80) acknowledges his paper on "Asymmetries and Mind-Body Perplexities" as a major influence, due to its deployment of the use-mention distinction. Importantly, Gunderson emphasises the link between leaving the subjective out of our descriptions of reality - which Loar treated as a purely semantic point – and subject deflationism, which is an ontological claim, and further strengthens the case, made in [4.5]-[4.5.2] that the use-mention distinction can help make sense of illusionism.

#### **[5.4] Gunderson on Stepping Outside of Ourselves**

Gunderson poses the mind-body problem quite differently from Chalmers. He focuses on selfhood at the expense of qualia, and in particular on the relationship between knowledge

and the possession of a perspective which Wimsatt drew in [4.5.2]. The critical metaphilosophical issue for a heuristic account of cognition is that if philosophy is a perspective in Wimsatt's sense, how could we know its limitations? This is particularly pertinent to first person foundationalist or 'armchair' conceptions of philosophy are in play, as they are almost inevitably on the subject of the mind-body problem. If Wimsatt is right then the first personal perspective, as a conceptually isolated tool for accessing ourselves, should be expected to overextend without our noticing, hence the conceivability of solipsism, idealism, panpsychism, etc. But as a major metaphilosophical motive for taking armchair intuitions seriously, the problem of consciousness thus encountered precludes getting a clear and objective view on whether to prefer antifoundationalist naturalism, as I argued in the introduction. Wimsatt's account, if correct, would prevent us from knowing whether or not it is correct. Gunderson tackles this head-on, starting with the idea (which Moevs defended on behalf of Dante in chapter 3) of the first-person perspective as the ultimate or final perspective by which all others may be understood, contextualised, or spatialized.

It is not so much that one boggles at conceiving of any aspect of his self, person, or consciousness being described in physicalistic terms; it is rather that one boggles at conceiving of every aspect being simultaneously so describable... A physicalistic (or otherwise monistic) account of the mind at the outset seems quite convincing so long as I consider anyone except myself. (Gunderson 1970, p274)

Consequently, the fact that science is done by human investigators who each encounter the world from a first person perspective turns out to be a major obstacle to completing a scientific theory of everything. This is why Wimsatt's claim that the possibility of spatially

objectifying the mind is an active hypothesis seems to be question-begging. Just pointing out that we could know a lot more about ourselves, just as we could know a lot more about the world around us, won't immediately point to an intelligible solution to the mind-body problem. *Could* we ever reach all the way to the bottom of things, to a full account of the conditions of possibility for one's own existence? Any remainder left out of the investigation will motivate identifying oneself with precisely that – with whatever is beyond our most developed physical concepts to conceive of.

Gunderson poses the issue as a kind of paradox: "Let us suppose that at any given time we can only investigate ourselves up to a certain point...where, say, some information-processing center in the brain, some subregion of the cortex, is unable to process information about its own information processing" (302). He continues:

Let us call this point P. Let us further suppose that we are able to investigate others beyond point P, to point BP. And let us assume that what we find out at BP is similar in kind to what we find at P. What might at first sight seem odd is that what we find out by advancing to BP in the case of others should in effect be an indirect explanation of the nature of whatever there is to our own minds which we seem utterly unable to investigate in the way in which we can investigate ourselves up to P. It is easy to assimilate the rest of our mental life to this seemingly intractable residue of self... [but] what we find when we advance to point BP in the case of others will intuitively seem to have nothing to do with the clarification of our own self and what we take to be its various aspects, namely thoughts, feelings, and sensations (302-3).

Like Lockwood (and Balog; see n.5), Gunderson thinks this paradox can be explained away as a consequence of the fact that the eye cannot see itself seeing. It would be absurd to

suggest that one's own eyes must have a special property that prevents them from being fully visible to oneself (298-9). And it would be implausible if the mirror somehow failed to disclose some property of one's eyes, given that everyone else's eyes are fully available to public scrutiny. Nevertheless, there does seem to be a problem, in that the idea that vision *as such* has limits to what it can disclose to each of us might be thought to cast doubt on whether we see anything quite as it is, and on whether anyone could know the nature of vision as such and hence, things as they are in themselves. But this would be a fallacy. It would be inferring from the fact that representation is necessary in order to perceive mind-independent reality, to the impossibility of ever perceiving reality: *we have eyes, therefore we can't see.*<sup>11 12</sup>

It might therefore seem as if needing to be *aware* of all the physical facts before knowing whether physicalism is true is asking too much. We don't need exhaustive self-representation to represent ourselves *as* exhaustively physical. But we would do in order to represent representation itself, if that is what it would take to convince us (remember that the 1<sup>st</sup> person is plausibly the 'master' representation to which all others refer). If so, it can be convincingly proved that doing so is logically impossible: Gunderson argues it would be formally equivalent to a self-embedded sentence, and this would be incoherent, for "no whole sentence can be one of its own constituents" (298).

---

<sup>11</sup> As Franklin (2002) phrases it in reference to David Stove's "worst argument in the world" (Stove 1991). Stove accuses all idealists of committing this simple fallacy, which he calls a "gem", whereby an interesting conclusion is derived from a tautology (Loar also phrases his objections to anti-physicalism in these terms). As Karl Ameriks (1990) has pointed out, this would be a "short argument" to idealism, where "reflection on the mere notion of representation, or on such very general features as the passivity or activity involved in representation, is what is meant to show that knowledge is restricted from any determination of things in themselves" (p. 63). Ameriks notes that Kant himself advanced a "long argument" for, among other things, the necessary ideality of time and space, which was not subject to the Stove's criticism.

<sup>12</sup> There does seem to be the possibility of empirical research revealing that perception cannot be direct in the way common sense might assume; indeed, Lockwood regards science to have done precisely that. But he stresses that a causal account of how representations relate to reality will tell us what representation is, rather than demonstrating that representation is impossible. Since that account will most likely refer to cognitive heuristics, and since Wimsatt's account of cognition may be thought to beg the question in favour of naturalism, I will leave open the sceptical possibility of our having no access to reality after all for the time being – Stove's Gem notwithstanding. The next section and the following chapter attempt a 'dialectical' solution to these difficulties.



Wimsatt would no doubt draw the conclusion that this is why we need robustness and second-person triangulation instead of foundational truths (because we can't reach the foundation). But simply revising what we mean by truth and representation could be seen as question-begging at this stage, in favour of naturalist epistemology and psychological explanations for metaphysical problems. Gunderson instead seeks to *dissolve* the paradox of going beyond point P. Perhaps even more absurd than seeking to see the seeing in one's own eyes, or constructing a self-embedding sentence, is the effort to interpret one's own speech acts as *needing* interpretation in the first place; that is, to perceive one's own utterances as they are perceived by others (287-8). Since my own utterances don't need interpretation, they present me with no properties requiring further explanation; I just say what I mean, and what I mean is *that*, the utterance I am saying. The metasemantic fact that my words mean anything has no particular content; there is no mysterious property of meaning inside or behind the words themselves.<sup>13</sup> So if from the first person we encounter no special mental properties, and third person description leaves nothing out, then where is the mystery? Logically, it seems there can be no mystery. In consequence, Gunderson proposes to understand all these cases – of eyes, sentences and utterances – in a “metaphysically neutral” way reminiscent of the Identity Theory's topic-neutral analysis (301). My utterances, for example, are nothing but the same *sorts of things* that require interpretation when uttered by others; “I am identical in kind with what I find other people to be” where by “what I find other people to be” is meant as they are (or might be) revealed to be on the basis of empirical investigation, etc” (*ibid*).

Gunderson's solution might seem to be overdetermined. If there is no special problem of meaning from the first person (or special content to phenomenal concepts, for that matter, since this is an early version of PCS), then there could be no appearance of

---

<sup>13</sup> This underscores Wimsatt's point that full understanding of a subject matter is necessarily demystifying.

asymmetry between self-ascriptions of intentional and phenomenal content and ascriptions to others. Conversely, if science's ability to demonstrate to one's own satisfaction whether physicalism is true were neither here nor there (301), because the satisfaction conditions are a priori absurd, then there would be no need to fall back on the topic-neutral analysis of consciousness (whose modern manifestation in PCS's claim that phenomenal concepts are bare demonstratives came in for such criticism in the previous chapter); we could have full-blown dual revelation, and still satisfy ourselves that physicalism is strongly necessary.

The previous chapter dealt a blow to both interpretations: phenomenal concepts are beefier than bare demonstratives, *and* dual revelation is incompatible with either physicalism or any other respectable metaphysics. Gunderson cannot pretend to be metaphysically neutral, anyway, and yet insist at the same time that third-person descriptions are closer to the truth than the first-person. Which means that "the fact that I cannot make myself an object for public inspection by me and still remain myself" (304) is not to be read as a tautology, but as a substantial, metaphysical claim that there is in fact no such self. Gunderson ponders the possibility of "language sans egocentric particulars (following Russell)" (*ibid*), which if coherent would save us from asymmetry, as well as "a question too crazy and deep for me to consider here... In the last analysis might physicalism and panpsychism turn out to be the same doctrine?" (309) Revisionist metaphysics like this would not even be desirable if the impossibility of fully representing oneself were in and of itself a solution to the hard problem.

As before, there seems to be an issue of how we could ever *know* one way or another whether or not there is a self. Gunderson does not seem to have advanced from the metaphilosophical dilemma with which I began this section, and which was previously posed in [4.5]. In particular, the idea that consciousness is metaphysically neutral seems just as question-begging as naturalism's denial of Cartesian foundationalist epistemology, despite

the fact that Gunderson's conclusion seems to follow logically from our not needing, first personally, to see ourselves seeing to know that we are seeing, or to interpret our utterances to know they have meaning. That is, Gunderson gives an argument that meets Descartes on his own turf – the first person perspective – and yet arrives at conclusions that offend intuition.

As a first approximation, perhaps Gunderson has gone wrong in assuming there is no room for doubt about one's meanings and experiences. The former is particularly easy to entertain, as I might wonder whether my utterance was a Freudian slip; but more general sceptical doubts about meaning seem to be conceivable as well. If Gunderson were right that meaning is strongly, indeed trivially, identical with whatever neuroscience comes up with, then it would be impossible to imagine empirical evidence failing to tell us what we meant by meaning. But it is perfectly sensible to doubt whether teleosemantics is an acceptable candidate for the meaning of 'meaning'. Similarly, not all reasons for doubting direct realism can be as fallacious as 'short' arguments for idealism (n.12, above). One might agree with Lockwood that directness furnishes no criteria to distinguish the true from the false, and so makes no gains on indirect realism. One might also agree that mechanistic explanation in the natural sciences rules out direct realism a posteriori. This gives at least some foothold to the worry that vision fails to disclose things as they are, and therefore that we are not *really* seeing according to common sense definitions of the term. With regards to phenomenal experience, the problem seems to be that we *cannot* doubt that there is more to experience than the neutral fact that we are experiencing anything in the first place: we really do seem to have qualia. Nevertheless, if Chalmers is right that our certainty is a matter of acquaintance with special properties, then this at least introduces the possibility of a gap between phenomenal judgements and experience. My zombie twin's doubts about its experience would be perfectly coherent, even if the same thoughts (identical down to the

last neuron) would be senseless in my own case. So there is something to be said for Gunderson's point that our Cartesian certainty does not hinge on knowledge of properties.

Similarly, there is surely something to the thought that, *if properly understood*, meaning and consciousness will seem unsuitable to form the basis of either sceptical doubts or metaphysical claims. In the case of meaning, the thought is that it is resistant to meaning scepticism because there is nothing there that one could coherently doubt. And perhaps supporting belief in special properties by appeal to introspection would be similarly senseless if we knew what introspection was. So perhaps Gunderson's argument is not overdetermined after all. Maybe our inability to fully represent ourselves means we're not privy to *why* our mental concepts are *actually* topic neutral; why the semantic meaning of our thoughts is the same as their having no meaning, and why panpsychism is the same as physicalism. Gunderson suggests this by claiming to have explained the grain problem as an illusion caused by thinking there must be non-gappy perspective in a gappy, perspectiveless brain. He notes that the closest likeness to seeing another person's perspective from the outside would be to display their location on a screen, *but without their body being in view* (307). Just as Wimsatt predicted, the impression of a purely subjective universe results from *lacking* information about our bodies, information that could only be accessed from other perspectives. But unlike Wimsatt, Gunderson makes his case on purely logical grounds (what meaning and experience must consist in even if it feels like there is something more). Paradoxically, the a priori impossibility of fully representing ourselves is the closest we can get to a view from nowhere from which mind/brain identity will just seem trivially necessary.<sup>14</sup> Only the psychological fact that we continue to believe that we can fully represent ourselves after learning that it is impossible needs explaining by reference to contingent, empirical facts such as the ubiquity of psychological heuristics.

---

<sup>14</sup> In an inversion of the unlimited self-knowledge Dante claimed in ch.3 to give a God's eye view on things.

There is no reason why one shouldn't read metaphysics into Gunderson's supposedly neutral analysis: Rather than reflection – prototypically, the act of Cartesian doubt – creating the illusion of substantial phenomenal insight, as Goff claimed, but which Raffman argued was impossible since we cannot posit Raffman qualia into existence, reflection for Gunderson momentarily reveals that there is nothing more to special subjective properties than appearances. We are unreflectively under the illusion that we enjoy phenomenal insight into non-conceptual properties, but upon reflection we gain rational certainty of non-phenomenal self-identity. Neither beefy phenomenal concepts nor bare demonstratives exhaust the *structure* of experience, as Raffman establishes, but thin concepts do exhaust the intrinsic *stuff* by which all possible experience is structured. However, to pursue this line of argument any further, it is important to re-establish a priori metaphysics as a genuine possibility, against the naturalists who seek to explain it away, and to avoid the pitfalls of pre-scientific Aristotelian metaphysics, in spite of there being a clear overlap between Gunderson's attempt to dissolve mind-body asymmetries and that which Moevs touted on behalf of Medieval mystics.



## Chapter 6:

# Neutral Monism Against the Qualiophiles

### [6.1] Wimsatt's Incoherence

Wimsatt summarises his metaphilosophy as one in which it is irrational to aspire to be perfectly rational (2007, p17). Nevertheless, we might wonder whether the ideal of perfect rationality is entirely out of play when we judge from an apparently ahistorical, impersonal perspective that it's usually more expedient to use heuristics. Wimsatt still thinks the extent to which methods err can be measured, which also means being less than sanguine about error as such. But given the story he is telling, misrepresentation is just a *necessary* feature of representation, and so not really an error or, for that matter, a feature that could be represented. If there is no general, all-purpose heuristic (as Wimsatt tells us repeatedly that there isn't), then from what position is he writing when he gives all-purpose advice? He tells us that "An adequate philosophy of science should have normative force" (26). But it's not clear how a metatheoretical reduction of reason to heuristic strategies could do this.

Wimsatt's approach seems as if it would have to be descriptive.

This is particularly pertinent when the model of rationality as a heuristic is used to draw *lessons* for future philosophy. Wimsatt tells us that extremely complex phenomena (prototypically, humans) will demand a multiperspectival analysis, not only as many scientific

models as we can throw at them, but the full range of literary, historical, political, and philosophical points of view as well – and even then we can expect these all to be exhausted before their subject matter has been entirely fathomed. The excess of a phenomenon to any means of conceptualising it is termed a “causal thicket”, and is predicted to cause controversy among the various points of view on it, since many of these will be incommensurable with one another. At this point, “the boundaries of perspectives begin to break down and it becomes more difficult to decide which perspective (or perspectives) a problem belongs to” (238), which can lead to “functional localization fallacies” (35). This is “such a general problem that it infects all areas of philosophy. What else is G. E. Moore's "Naturalistic fallacy" (if it is one) but a kind of mis-localization claim?” (*ibid*) So he thinks he can legislate on when a philosophical problem is illusory. But he gives no general criteria for when “violations of a rule may indicate our own failures rather than problems with the rule” (16), which would also apply to when we should take a contradiction as evidence that we’re using the wrong heuristic.

In turn, this means the problem of when a heuristic has been applied or misapplied has no general-purpose answer. Wimsatt is keen to enlist philosophers to the cause of settling these sorts of boundary disputes (*ibid*), but the broader, overarching perspective philosophers normally bring to the table, whereby relativism’s relative plausibility is brought, ultimately, into disrepute, is exactly what Wimsatt’s approach rules out. He thinks relative truth is *ultimately* all we have.<sup>1</sup> So there is still the problem of how we know a contradiction is avoidable by adjusting our methodology versus being an inevitable upshot of it. What if we’ve tried everything and still can’t rid ourselves of contradiction, but there’s no alternative method we can use? And how could we rid Wimsatt’s approach of contradictions?

---

<sup>1</sup> “Is this attacking logic? Logic doesn't break down by being approximately right. It is a reflection of the standing of logic in our scheme that we wouldn't think of violations in this way.” (33) Wimsatt is overtly relativist here.



In keeping with his analysis of subjective experience as a coherent object of study only insofar as it can be situated in an objective context, Wimsatt attacks the very notion of any *one* perspective being in a position to teach us anything:

suppose one did have only one means of access to a given quantity. Without another means of access, even if this means of access were not made definitional, statements about the value of that variable would not be independently testable. Effectively, they would be as if defined by that means of access, and since the variable was not connected to the theory in any other way, it would be an unobservable, a fifth wheel: anything it could do would be done more directly by its operational variable. It is, then...a bridge that leads to nowhere.  
(Wimsatt 2007, p54)

Yet our means of access to Wimsatt's metatheoretical claims is singular: our capacity for rational thought. There can only be one metatruth about what kinds of truth are available. He tells us that he wants "criteria for what is real that are decidedly local – which are the kinds of criteria used by working scientists in deciding whether results are real or artifactual, trustworthy or untrustworthy, objective or subjective" (195), as if these are the only perspectives worth catering to. He warns that "traditional foundationalism and ontic fundamentalism are in trouble. They will survive, if at all, as a local kind of problem-solving technique of significant but limited usefulness" (*ibid*). But this last qualification puts the cart before the horse: foundationalism's usefulness is limited precisely to the kinds of metaphilosophical issues Wimsatt raises, which will be of no interest to scientists.

Those same scientists may wish to ignore philosophical attacks on the very idea of reducing rationality to a local, contingent, context-relative bag of heuristics. After all, there does not seem to be any barrier to such a descriptive fact about us being true, and indeed it

is difficult to imagine the naturalisation of thought yielding any other conclusion. So it will be a matter of seeking an absolute point of view within the framework Wimsatt presents us. Luckily, we do not have to look far. On the issue of robustness as a criterion of reality or truth, Wimsatt notes that it “can only make the probability of failure decline-though it can get very small, it does not go to zero. This criterion does not give certainty. *Nothing does*” (196, emphasis added). Indeed, the assumption that “our reasoning processes must be boundedly rational...will *never* become obsolete” (320-1, emphasis added). And yet he warns us not to “accept self-trivializing, extreme, brittle, and fashionable theories that prove too much too easily; for example, that the results of scientific investigations are (just) socially determined. Beware of all varieties of “nothing-but-isms” ... just as we should be suspicious of “single-factor” theories of anything” (341). So how to judge the claim that all thought consists of nothing but heuristics?

Wimsatt’s claims are apt to being rendered in lofty, paradoxical sounding terms, such as that philosophy is doomed to performative contradiction, and thus failure, in trying to comprehend the biological foundations of its own enterprise, and to succeed only insofar as inconsistency is tolerated. Claims that would appear to have been delivered from on high by scientific experts who answer to no-one, least of all philosophers. Putting it this way, however, is question-begging and obstructive, since it forecloses the (obviously very real) possibility of being found guilty of (perfectly ordinary and uninteresting) hypocrisy and incoherence, a charge which scientists are as much expected to answer rationally as are philosophers. Of course there are many contexts in which thinking clearly about a subject matter runs counter to one’s intuitions, e.g. QM; here laypeople and scientifically uninformed philosophers are better off suspending their critical faculties and deferring to scientists. But questions of how to interpret a philosophical text aren’t apt to be turned over to a body of experts in quite the same way, since there is no determinate, mind-independent

reality there for institutional error-filters to whittle out. Insisting that philosophy be subordinated to scientific criteria of rationality, as Ladyman and Ross (2007, p29) argue, is simply the obverse foundationalist error to that of making oneself the sole arbiter of sense-making and of what ought to pass as a charitable interpretation (philosophical delusions of impotence and omnipotence, respectively). And asking loaded questions, such as insisting metaphysicians specify an ecological context or 'niche' in which their research could be seen to have instrumental value in natural scientific-sounding terms, is guilty of both errors simultaneously: deferring to science as an esoteric, fixed body of doctrine capable of determining, in advance, what questions are worth asking and what their answers must look like, while identifying problems amenable to its supposed methods via nothing more than an interpretive squint. But these are precisely the questions which philosophers can expect to be battered with if the image of cognition presented in sections [3.5] and [3.6] came to be unconditionally accepted.

Wimsatt himself is clearly sensitive to this problem, and seeks to distance his position from "postmodern" relativism (148) on precisely the grounds that it paralyses critical thinking. It needs to be remembered that Wimsatt's claims ('it's all heuristics'), if true, are themselves heuristic strategies. The aim is for readers to know their way around the very complex worlds of psychology and science a little better than before; these fields comprise the 'niche' of Wimsatt's book. So diagnosing philosophical problems, or descriptive properties of philosophical discourse, as nothing but misapplied heuristics, in a theoretically and interpretatively unconstrained manner, would seem to be strictly beyond the scope of the book. The 'heuristic' heuristic cannot be applied willy-nilly; there are contexts in which resorting to psychological explanation of one's own and one's interlocutors' claims is question-begging and obstructive, since it undermines the virtues of logical deductive thinking and humility. The 'second person' of scientific error-filtering is, as Wimsatt himself

emphasises, continuous with the common-sense giving and asking for reasons of the cultural conversation, a context in which critical and self-critical thought are highly valued. However, that continuity can be understood in different ways. If it means that psychology cannot be fully naturalised without undermining rationality, then Wimsatt would be a scientific anti-realist, but he shows no evidence of this, and maintains his position is compatible with realism (149). Indeed he insists that "Science and technology are the most productive systematic sources of knowledge production we know...they are the best we have" (339).

Wimsatt may also insist "that it is in one's interest to come to understand differences, and then to resolve them" (239), and warn that "eliminativists should beware: you don't make friends with the natives (folk) by denying their legitimacy (psychology), and you can't tell what's in the territory without a native guide. You can play imperialist without heeding these warnings, but it usually requires more resources" (240). But he is only as accommodating to other perspectives as his rather half-hearted relativism permits, and we have already seen that some perspectives – the mechanistic – are more equal for him than others. At most, Wimsatt's liberal naturalism will permit a wider range of objects than, say, Quine or Sider would have, and then only insofar as they have "causal effects" (Wimsatt 148). He sees his position as a kind of "classical mechanistic materialism" (380). This is highly revisionary, and Wimsatt makes no bones about its implications: he hopes to explain away all mental essentialisms as misinterpretations of causal thickets (190-2). Boulter and Wilkes had the same problem rescuing 'common sense', which they in fact reconceived as a set of reliable heuristics for tracking mind-independent causal powers.<sup>2</sup> Common sense beliefs are

---

<sup>2</sup> I think Papineau can be read in this way as well, and therefore spared the embarrassment of committing the sympathetic fallacy. His realism about mental states is highly qualified, as I discuss in more detail below, and commits itself as little as possible to intrinsic properties. Goff may still be right that for Papineau the Spirit of the Woodland will be *no more or less* real than mental properties, whatever that amounts to, but this may turn out to be harmless. Goff thinks of strong necessities as offering a (bad) analysis of properties, licensing unrestricted predication of mentality, or normativity, to objective states of affairs. By contrast, Heil would regard Goff's portrayal of PCS (whatever its independent merits) as presupposing the Canberra Plan's approach to concepts. Heil argued against the latter that descriptive facts about what we find interesting restrict

also normally thought to be mutually supporting; perceiving other minds presupposes realism about the external world, folk psychology, and plausibly free will and morality as well. But these all received separate treatment by Boulter, since they must each be validated or rejected by evolutionary psychology, and even one of these being indefensible would destabilise how we understand the others. It seems fairly clear that the problem of consciousness will be a major sticking point for this sort of redescription, as Wilkes admits, and so the hope of a commonsensical naturalism looks dim.

It looks like any solution to the hard problem is going to be counterintuitive in one way or another.<sup>3</sup> Rather than following Goff and seeking to bypass mechanistic explanations altogether, however, I suggest philosophers are better off bolstering rather than downplaying the imperialist ambitions clearly evident in Wimsatt's philosophy of mind by putting it on a secure footing, one which restores the dignity of traditional foundationalist philosophy by reconciling it with modern science. As I have repeatedly emphasised, doing so began with Descartes.

## **[6.2] Rationalist Metaphysics and 'Stress Free' Modal Dualism**

---

predication of (e.g.) to-be-doneness, since values are inextricable from the practices in which they emerge (a Wittgensteinian point Heil shares with Wilkes), which is also how macroscopic objects emerge.

Once ontology shifts to descriptions of practices, then even Goff's facetious example may have more merit than he supposes. Animism is a useful means of interfacing ourselves with the natural environment, as the ecologist David Abram (1997) argues at length. For Abram the continuity between language and causation explains the possibility of communication with animals and even inanimate objects, a 'dialogue' he argues is exhibited by all indigenous cultures still living off the land. Teleosemantics is compatible with all of this so long as it is understood as shorthand for mechanistic interactions. Evolution can explain why, as social animals, we have more brain power adapted to taking the intentional stance and remembering narratives. So it does not seem unreasonable that reading the natural world in these terms will exploit a greater share of our intelligence, and greatly increase one's chances of survival, just as visual areas of the brain are co-opted in the congenitally blind into providing greater tactile and audio perception than would otherwise be available. All of this can be explained from an intellectual, mechanistic point of view despite the fact that, as Abram argues, this perspective impoverishes our relation to the environment.

<sup>3</sup> See Schwitzgebel (2014).

For Descartes the innate ideas of thought, extension and God make scientific explanation possible by imbuing what appears to us with rationality. Perception of physical continuity is due to the innate idea of extension which we infer in objects rather than perceive in them, as in Descartes' famous example of how we know the same wax can be both a solid and a liquid. Quite apart from the specific concepts we use to measure reality, Descartes' overarching position is that knowledge of things in themselves is an exercise of theoretical imagination, not sensation. How we think, or represent, the world, is now more a scientific question than it was in Descartes' time, and so any account of what theoretical or symbolic imagination consists in will be constrained by developments in the sciences. As a first approximation, though, these types of imagination do not make essential reference to qualitative properties; or rather, when they do refer to qualities, they represent them in a non-qualitative way, e.g., as representations-as-of-phenomenal-green; this being the illusionist upshot of PCS's use-mention distinction.

One might object that PCS cannot be too closely allied with rationalism if it ultimately appeals to 'non-logical' or 'strong' necessities, which were invoked in order to reconcile qualia with physicalism. And making physicalism the upshot of the triumph of symbolic imagination makes it look like it presupposes logical supervenience, which various naturalists attacked in previous chapters; naturalists who identified themselves Aristotelians rather than Cartesians. In reply, the kind of rationalism I am considering here is Patterson's Descartes, not what she refers to as the classical 'Cartesian' reading which attributes Descartes with a need for perfect certainty. Boulter's historicist attack on logical conceivability as such was a red herring, as the (PS) was above all about perfect determinacy and logical conceivability as criteria of *modal* truths, whereas I am interested in the *utility* of rational reflection and pure understanding. Theoretical and symbolic imagination could both be broadly construed as comprising any methods which do not pretend to perceptually

imagine their subject matter, and so could accommodate a posteriori falsification and second-person triangulation of evidence as well as a priori proofs of coherence. 'Mentioning' things rather than 'using' them is a broad category which only rules out first-person sources of evidence, just as the main thrust of Cartesian metaphysics, according to Patterson, is that subjective impressions must be set aside in order to understand things in themselves (Patterson 2000, p99-100).

Once again, by 'mentioning' a property rather than using it one is able to situate that property within a wider context; only when a property is not used can it be represented objectively, as it is. *The sceptical distance of the mind from appearances is a prerequisite for achieving a comprehensive understanding of those appearances.* This aspect of Cartesian metaphysics has certainly survived, even if the clarity and distinctness of Descartes' three innate ideas has been eroded by counterintuitive scientific discoveries, such as the possibility of machine intelligence, and the substitution of mathematical objects for substance in quantum physics. While science no longer reduces things to particles, mechanistic explanation is still reductionist, in the sense of reducing things to parts and causal relations, as both Bechtel (2007, p21-2) and Wimsatt (2007, p377) emphasise. So Boulter and Wilkes protest too much when they deny science resembles the Cartesian "geometrical method" any more. The content of scientific theories might be messier and more ambiguous, but their form still emphasises primary over secondary qualities; "in modern jargon, *the primary qualities are robust and the secondary qualities are not*" (Wimsatt, 2007, p198).

This is where scepticism comes into play, as it can act as a corrective to received wisdom. Paradoxically, it is only upon reflection that we rediscover that thought is an unreflective pattern of habit. Patterson's reading of Cartesian scepticism as a methodological corrective to received wisdom is consistent with fallibilist naturalism: the

standpoint of objectifying, disembodied rationality, discovered after radical doubt, provides a norm for getting certain kinds of robust results; e.g., situating things in space reassures us of the unity of nature. Even the disappearance of matter from contemporary physics (see [3.4.1]) was made possible by applying the same norms which reduced everything to objects in the first place. From within a purely methodological naturalist framework, Cartesian scepticism shows its worth by putting us in a mindset that can engage with mechanistic reductionism in the first place. It therefore serves as a master heuristic for entering into Wimsatt's metapsychology, justifying its imperialist ambitions to naturalise the mind.

There is more to Descartes than just a methodology, however. His aim was not simply to understand the world better, but to understand oneself – not just heuristically, but metaphysically. In this respect there is a close connection between Descartes and Plato:

for Plato, knowledge is won by freeing the intellect from the body and the illusions of the senses, so that mind may perceive or remember the ideas of things in its own being from before its incarnation, *despite* the world. For Aristotle the world *is* intelligibility and being; knowledge is the capacity of the rational soul to become the world. (Moevs 2005, p56)

As I argued in chapter 3, the innovation of materialism was to invert the Aristotelian picture and put unintelligibility first: understanding, so conceived, is a recently-evolved phenomenon, derivative of prime matter, which Descartes was the first modern philosopher to postulate as an independently existing substance. Wilkes and Boulter's professed Aristotelianism missed this distinction by focusing on the brain's capacity to form representations, which they liken to *nous's* capacity to 'become' the forms it contemplates. Of course there is a limited sense in which brain processes are informationally identical (really, just relevantly similar in some aspects) with structures and dynamics in the world.



The brain *could* be likened to a subtle wax on which forms are impressed. But making intentionality subordinate to the compositional possibilities of matter inverts Aristotle's prioritisation of form: it portrays knowledge as a type of matter, the upshot of processes that are neither rational nor all-encompassing. These philosophers' functionalism is not ontologically serious enough about intentionality to qualify as properly Aristotelian; on the contrary, it is importantly similar to Descartes' causal account of perception, of which teleosemantics is a descendant.

Descartes went further than causal accounts of perception, however, and argued that the body was a hindrance to understanding (Patterson 2000, p99-100), rather than the vehicle of understanding: Cartesian metaphysics motivates withdrawal from the world in much the same way as Platonic metaphysics. And unlike Aristotle's 'chain of being', in which the highest 'spheres' were still continuous with common-sense reality, the current scientific paradigm is only available to an intellect that is *dismissive* of much that was once considered obvious, including the idea of purposes or final causes in nature that the mystic's ascent beyond the spheres was supposed to progressively reveal.<sup>4</sup> Matter for modern physics exists independently from teleology and universals insofar as these figure in everyday life, and is unintelligible to sensory imagination. As argued in [4.5.1], being unintelligible to sensory imagination is a plausible definition of strong necessities that does not lapse into irrationalism: it offers a "stress-free" modal dualism, albeit of a different kind than Goff intended (as discussed in [4.4], and especially n.37 of that chapter). Descartes wrote that the laypeople of his age "are so used to thinking of things solely by imagining them (a way of thinking specially suited to material things) that whatever is unimaginable seems to them unintelligible" (quoted in Patterson 2000, p80). Goff's definition of intelligibility, defined as

---

<sup>4</sup> Bechtel (2007), p10-11 argues that physical laws are a throwback to teleological forms of explanation prevalent under Aristotelian metaphysics. By contrast, mechanistic explanations focus on what outcomes can be expected as a consequence of how a thing has been contingently put together.

the possession of transparent concepts whereby we completely understand what we're conceiving of, is comparatively loose, indeed arguably circular, as he does not clarify by what *criteria* we could judge ourselves to fully understand something. Cicero being Tully might, in this regard, be a bad example of what troubles Goff and Chalmers about modal dualism.

Papineau and Hill's decidedly Cartesian distinction between intuitive and theoretical beliefs is where the real 'stress' of modal dualism lies; for it suggests that our understanding of consciousness will be no more intuitive than our grasp of higher-dimensional objects.

Chalmers' complaints (2010b, p181-2) that psychological explanations for our intuitions are psychologistic rather than logical can therefore be met by replying that theoretical or symbolic imagination is stringent enough to rule out thoughts about square circles and so on.

Gunderson similarly registers the power of non-presentational imagination, which 'mentions' rather than 'uses' its objects of reference, when he concludes that "If any kind of dualism is to show its mettle, it must now do so from a third-person standpoint" (Gunderson 302). He presumably has in mind sensory imagination and empiricism, but even neuroscience (let alone fundamental physics) deals with imperceptible entities. Given that there is nothing more thoroughly third-personal than symbolic imagination which is insensitive to how things would appear to us, Gunderson may just as well be read as posing the challenge of coming up with a thought experiment that doesn't make *any* assumptions about the reliability of sensory experience. In this vein, Hill considers whether the *conceivability* of mind-brain separation is refutable by the separation of faculties. He points out that "it is in principle possible to use the faculty of conception to construct representations that are largely or entirely without a qualitative dimension." (73)

The challenge is then as follows: If dualism seems intuitively plausible even in acts of imagination that don't themselves use (as opposed to mention) phenomenal concepts, then

anti-physicalism would have a non-question-begging argument on its side. At this point it looks as if the extreme abstraction of modal rationalism works in its favour: Chalmers' zombie argument plausibly 'mentions' the notion of 'all physical facts' without using them, that is, without actually having them all to hand. But if I cannot intuitively imagine having all the physical facts, then I cannot have any intuitions about whether they allow for zombies or not. To be sound, the argument would need a non-qualitative concept of qualia as well, and that concept will just be: 'an experience of "*that*" type', where 'that' is just whatever mind-independent property the concept pain picks out. Which in turn makes no reference to properties that would resist mechanical reduction; similarly, there is no appearance of contingent mind-brain relations under a physico-mathematical description of the brain. The appearance of contingency comes from giving more credit than is due to the powers of sympathetic and perceptual imagination to give insights into the true nature of things.

For Descartes, the mind is strictly separate from sensations, and more certain than them. To the extent that we know we have sensations, it is because we know we have a body, but we only know *that* because (according to Descartes) we have the innate idea of God. Of course without God's safeguard things would still *seem* a certain way, but then appearances would be deceptive, and their ontological status unresolved: only the fact that one is thinking (and therefore existing), as opposed to feeling, could then provide a platform for metaphysical speculation. *That* one is thinking is certain; as to the reality of that about which one is having thoughts (*that* to which phenomenal concepts refer), this was for Descartes as much a scientific question as it is for us: for all we know, sensations may be physical properties. Whereas the *cogito* is discovered entirely via non-qualitative intellection.

The previous chapter argued at length that Raffman's account of perception could ultimately countenance bare demonstrative concepts, in spite of Raffman's protestations, as

the only honest form of phenomenal concepts: a zero degree or zero qualia version of them. Raffman argued that we have no subjective access to phenomenal properties, but only cognitively impenetrable causal relations with physical properties which escape our introspective heuristics. Hence, 'that' can be admitted as an expression of ignorance towards what phenomenal experience is.<sup>5</sup> However, it can also be thought of as directly conceiving a lower level of experience than that of ordinary empirical perception (the kind which represents mind-independent objects as possessing phenomenal properties). I suggest that experience ought to be construed in itself as a non-conceptual, non-qualitative 'something', and that it is this which is resistant to Cartesian doubt, as well as granting access to the kind of nature which could explain the success of science. The empty certainty of the cogito, the thought that one is thinking, should be reconceived as the certain existence of featureless substance which is ambiguously mental and physical: a neutral stuff whose only intrinsic property is to be self-identical.<sup>6</sup> Contrary to Goff's portrayal of it, the demonstrative 'that' may indeed give us phenomena insight, phenomenal certainty, and rational certainty, about the presence of zero qualia. Several philosophers have already anticipated such a view, and help spell out its implications.

### **[6.3] Translucent Consciousness and Neutral Monism**

---

<sup>5</sup> Similarly, illusionism doesn't challenge the reality of appearances, or deny that the appearance is the reality, so much as question whether we know the reality in which appearances consist. This is why it overlaps with PCS, wherein physical properties can have nonphysical modes of presentation. (Though for Raffman and Wimsatt it is not so much a case of phenomenal modes of presentation as an occlusion of facts which would allow us to deduce that the appearances are physical).

<sup>6</sup> To develop the previous footnote: Illusionism about the present moment can be distinguished from the indisputability of the 'there is'. This can be compared with the Aristotelian mysticism surveyed in chapter 3: there, all of creation was present at every moment, albeit in innumerable complementary ways. So the fact of presence needn't indicate perfect knowledge of what is present (what kinds of things are present) or what is past or future (i.e. it doesn't require our being able to infallibly distinguish between memory, occurrent experiences, and anticipations).

Why does consciousness's physicality seem contingent? Papineau argues (1993, p119) that the best PCS can do is show we cannot rule out that consciousness is identical to this or that brain state/function. We can't actually rule *in* any particular answer that would tell us which state or function it might be. However, Papineau also says "that physicalists should simply reject this question" (119-120), on the basis that "Being conscious isn't something over and above having A, it just *is* having A" (*ibid*). This seems to be just a defence of strong necessities. Indeed, he says it is "quite consistent with holding that consciousness is a real property which distinguishes some kinds of systems from others" (121-2). But Papineau then undermines this defence of the identity theory: "I shall argue that questions of consciousness may not only be vague, but quite *arbitrary*" (123). His grounds for this view were touched upon earlier at the end of [2.2.2], where he raised doubts about whether self-scanning mechanisms could account for the contents of consciousness. All of which is the upshot of his view that physicalism "denies that consciousness is some kind of extra inner light, some further non-physical property which exists over and above any physicalistically specifiable property" (121-2). Since physical complexity does not normally come in neat packages, Papineau "reject[s] the intuition that there is a sharp line between conscious and non-conscious states" (124).

At this stage Papineau is still treading ground that Loar would later go over in his discussion of the possibility that phenomenal concepts fail to pick out determinate properties. Like Loar, Papineau is apparently torn between following Wittgenstein's example and insisting that such doubts are senseless, and crediting them on the basis that experience exists independently of one's capacity for conscious self-reflection (see [2.2.1]). He allows that self-monitoring theories of consciousness might help individuate the property of first-person accessibility which underlies the heterogeneous mix of pains, moods, thoughts, perceptions, etc. But he also wishes to credit the intuition that animals and babies, which

lack the capacity for introspection, have experiences. Yet unlike Loar, Papineau takes the dilemma in a much more illusionist direction.

As briefly discussed in [2.2.2], Papineau raises the dilemma in a discussion of ‘mad pain’ and ‘Martian pain’. The former is a case of pain realized by the same processes (c-fibres, etc), but with a different functional role; it is caused “by moderate exercise on an empty stomach; and it doesn't cause [the madman] to writhe or try to alter the state, but rather to snap his fingers and think of mathematics” (131). Martians, by contrast, are supposed to have the same reactions to pain as we do, but different realizers (d-fibres). In the paper by David Lewis which first introduced these creatures, Lewis (1980) concluded with a compromise, in which pain is realized by its functional role for normal members of different species, but picked out by its physiological realizers *within* species (this is in-keeping with the Canberra Plan’s bureaucratic approach to philosophical problem-solving, in which ad-hoc compromises are to be accepted). Papineau objects that by rendering the presence of pain relative to what qualifies as normal physiology for a species, Lewis has left out the possibility of a human-Martian hybrid with d-fibres rather than c-fibres. Without some other objective arbiter besides function and physiology for whether the hybrid can feel pain (or perhaps more specifically, our type of pain), we are forced to suspend judgement.<sup>7</sup>

Goff raises similar objections to Lewis (Goff 2014b, p6-7) in order to motivate phenomenal revelation; we just *know* there is a fact of the matter as to what we are experiencing, independently of any ambiguities in the neural correlates of consciousness. Papineau of course takes the opposite line.

Surely the madman's experience is either like *that*, or it's not. What could be simpler?

---

<sup>7</sup> The example is perhaps more vivid in the imagined case of mad or Martian colour, since colour experience is much richer. Pain that is painful in a different way, due to different physiology, might be conceivable; red that is red in a different way for Martians would be better thought of as just a different colour.

I don't think it is that simple. The reason it seems simple is that we naturally suppose that, when we have (or imagine) a visual experience, we switch on an inner light. And so all we need to do is compare that shade of inner light with the shade illuminating the madman's mind. But there isn't any such inner light. There are just the physical and structural features of the relevant brains, some of which we share with the madman, and some of which we don't. So our conviction that either the madman must feel the same or feel different is based on a false picture. (Papineau 1993, p137)

It is at this stage that Papineau introduces his Wittgensteinian point about the time now just being the same as the time on the sun. “Here” and “now” are certain to refer when used by subjects with a point of view; but it is an ‘empty’ certainty. More importantly, Papineau is quite clear that this type of certainty has nothing to do with certainty about *what it is* in our experience that we are picking out with ‘this’ or ‘that’. On the contrary, his goal is to undermine the *meaningfulness* of doubts that physicalism leaves out some determinate phenomenal content: transforming into a pain-madman and back would be like being a frog immersed in increasingly less temperate water – one’s basis for noticing that *that* had changed would be undermined. This is different from Balog’s appeal to dual revelation as a way of *preserving* our intuitions about consciousness without discarding physicalism.<sup>8</sup>

---

<sup>8</sup> Balog agrees with Carruthers and Veillet (2007) that the epistemic situation phenomenal concepts put us in can be fully understood third-personally. She agrees with Chalmers that defending physicalism by insisting that phenomenal concepts be explained physicalistically is circular, but argues that the “Cartesian circle” of taking the first person perspective as one’s epistemological starting point is no less circular (Balog 2012, p19). She concludes:

“In the face of this kind of standoff one can be tempted to the view that there is no fact of the matter (i.e., the relevant statements are neither true nor false) or that the dispute is terminological... But we have seen that the conceivability of physicalism has consequences for ontology—for whether physicalism or anti-physicalism is true—and this doesn’t seem to be terminological.” (*ibid*, p20)

It is an irony of Papineau’s position that the dispute ultimately is terminological after all.

I am not suggesting that *how it is for* the colour madman will depend on how we classify his experience. Of course it won't. My claim is only that it is indeterminate whether the madman's experience is the *same kind of experience* as our experience of red... It's just that the notion of sameness of colour experience breaks down when we come to such cases.

(Papineau 1993, p136-7)

Papineau draws the same conclusion with regards to animals that lack a capacity for introspection or phenomenal concepts. So what he is saying is that how it is for most animals is *determinately indeterminate* with respect to what it's like to be them (an ambiguous *kind* of consciousness), rather than indeterminate as to whether there is anything (ambiguous or not) that it is like to be them.<sup>9</sup> In the former case, there is something even creatures without concepts can be certain of, even if it is both inexpressible and not, in any case, captured in descriptions of the contents of experience. In the latter case there would be no such certainty. This feeds into Papineau's residual sympathies for panpsychism. He can allow that the sheer fact of having experience is known immediately and may be ubiquitous, but that specifying what experience *is*, let alone what it is *of*, escapes us.

How do we know Papineau has sympathies for panpsychism? Goff relates the following anecdote:

It is an underemphasised implication of [Papineau's views on indeterminacy] (I have confirmed with Papineau in conversation that he embraces this implication), that there is no

---

<sup>9</sup> I think this saves Papineau from Raffman's critique of PCS. Raffman qualia, phenomenological overflow, and qualia occurring independently of self-monitoring mechanisms are all ambiguously conscious; and Papineau accommodates that. Papineau argues that the class of states to which mental concepts may be applied is too complex and heterogeneous for us to form criteria for the correct application of mental concepts, outside of arbitrarily limited contexts (e.g. by restricting the attribution of consciousness to just humans). Raffman likewise thinks there is no "specific complexity" grouped by our colour concepts. And Raffman could still end up with ambiguity about the nature and extent of animal consciousness. For example, Metzinger speculates that animals only have Raffman qualia, since they lack even a low-resolution capacity for re-identification (Metzinger 2003, p73).



fact of the matter as to whether or not panpsychism is true, just as there is no fact of the matter as to whether I am tall. On one sharpening of consciousness, the table and the pillar of salt are conscious, on another sharpening they are not. It is ironic that Papineau's denial of transparency, which allows him to escape the argument for panpsychism given in this paper, gets him in the end to panpsychism (at least on one legitimate sharpening of consciousness). (Goff 2014a, n.10, p82)

But of course the irony Goff notices is not a score for his position; quite the opposite. It is precisely because Papineau believes concepts such as attention and consciousness are vague that they are compatible with physicalism. In a recent exchange between Daniel Dennett and Papineau in the *Times Literary Supplement*,<sup>10</sup> Dennett goes so far as to argue that Papineau's ambiguous stance makes him a fellow illusionist.

On the face of it, it seems as if one must have to choose between either thoroughgoing physicalism or principled panpsychism. But the proximity between the two positions has long been recognised. Herbert Feigl, for example, argued for an early variant of the physicalist identity theory of mind; yet he was not altogether dismissive of panpsychism. He writes that "On the one hand "the identity theory regards sentience... and other [unexperienced] qualities...as the basic reality." He seeks to avoid "the unwarranted panpsychistic generalization." Yet, "one is tempted, with the panpsychists, to assume some unknown-by-acquaintance qualities quite cognate with those actually experienced" (Feigl 1958, pp474–475). And as Gordon Globus relates, Feigl once remarked that "If you give me a couple of martinis, a good dinner, and a couple of after-dinner drinks, I would admit that I am strongly tempted toward (a rather watered-down, innocuous) panpsychism" (Globus 1976, p320). What could this watered-down panpsychism amount to? A 1972 paper by Globus has the answer.

---

<sup>10</sup> "Papineau vs Dennett: a philosophical dispute" (August 4, no.5966, 2017).

#### [6.4] Globus on Pure Consciousness

After recapitulating Russell’s point that neuroscientists only directly perceive the inside of their own brains (Globus 1972, p292), Globus draws the following distinctions (*ibid*):

	STATUS WITH RESPECT TO TRANSFORMATION BOUNDARY	
	<i>Distal and Inferred</i>	<i>Proximal and Directly Given</i>
<b>Events-in-themselves</b>	<b>(1)</b> Nonneural events not embodied, e.g. <u>onta</u> [“elementary particles”]	<b>(4)</b> Neural events representing nothing, i.e. consciousness <u>per se</u>
<b>First Order Event-Representations for S</b>	<b>(2)</b> Nonneural events embodied, e.g. a <u>physical</u> line	<b>(5)</b> Neural events representing nonneural events, i.e. contents of consciousness, e.g. a phenomenal <u>line</u>
<b>Second Order Event-Representation for E</b>	<b>(3)</b> Neural <u>events</u> embodied in S’s brain	<b>(6)</b> Neural events representing neural events, i.e. contents of consciousness, e.g. phenomenal <u>brain</u>

In this example, we are to imagine *S* perceiving (5) a physical line (2), while *E* – a neuroscientist – observes (6) both *S*’s observation of the line as a series of events in *S*’s brain (3), and the line which is in front of them. This leaves (1), things in themselves as understood – inferred – by our best scientific theories, which act as a mediator between the apparently very different physical properties of (2) and (5), as well as the experiential differences between (5) and (6), despite the informational or structural properties of the physical line being somehow contained in all these representations. Like Lockwood, Globus thinks the major point of divergence between distal or ‘objective’ objects as we perceive them, and the

proximal or subjective experiences of which we are directly aware, is the spatiotemporal nature of the former, and like Lockwood regards this as an illusion: quantum physical objects are mathematically described 'events', not three-dimensional things.

*E* can infer that there is a real brain occupying space out there beyond his transformation boundary, but his particular conception may be only a function of his "Euclidean Brain" which represents warping in time-space curvature in a special way... The only distinction between what we confidently label a physical event, e.g. a red hot stove, vs. a phenomenal event, e.g. pain, is whether or not the brain has coded space-occupying information. (294)

So the appearance of a distinction between physical (extended) and phenomenal (nonextended) objects depends on which properties of fundamental reality are extracted by the brain; *and* phenomenal experience is closer to things in themselves insofar as they are not spatially extended.

There is nothing so far with which a staunch anti-physicalist such as Goff could disagree. However, Globus then introduces another distinction, between (5) phenomenal neural events-in-themselves representing distal objects and (4) neural events-in-themselves, *simpliciter*.

Now, the question may be asked, if phenomenal experience is identical with neural events, then what phenomenal experience is identical with the chronotopic patterning of literally billions of neural events...which do not represent the world distal to the transformation boundary but instead are intrinsic activities of the brain apparently involved with processing input stimuli and producing output behavior? (297)

Consciousness 'per se', as Globus puts it, is not conscious *of* anything in particular. As he admits, "this is a very difficult state to maintain more than transiently. (Meditators may train for years to attain the state of "supreme consciousness" wherein for extended periods there is no awareness of the physical world, the body or even the self, but simply pure consciousness" (298). This type of consciousness is "supreme", because it gets us closest to the properties of neurons when they are just *being neurons*, as opposed to busy inferring the properties of distal objects. And, furthermore, it gets us closest to a direct grasp of 'onta', or 'events', i.e. things in themselves *prior to representation*. "For example, Margenau ... suggests that location of onta may be added by observation, i.e. " ... that the electron *is* where it is measured, that it may be nowhere when it is not measured, that a measurement, properly contrived, may cause it to appear somewhere" (Margenau 1954, p. 10)" (299). Meanwhile, "Although it is most difficult to describe consciousness per se, it would seem to be event-like, a happening. It has no dimensions or location" (*ibid*).

Interestingly, Wimsatt (2007, p379-80) appeals to this paper as a potential solution to the mind-body problem, alongside Gunderson (1970). Despite repeatedly voicing sympathies for a use-mention solution, Wimsatt concedes here that the mental cannot and need not be fully spatialized, i.e. represented from a third person perspective, in order to qualify as physical. It is ironic that, though heuristics seem to rule out accessing things in themselves (as Frankish argues), Wimsatt ultimately appeals to philosophers sympathetic to panpsychism, whereby consciousness is the unrepresentable root of the physical itself.

On the other hand, Wimsatt's insistence that Globus's position "ought to be of no comfort" (*ibid*) to anti-physicalists might be somewhat borne out by Globus's own failure to fully pursue the radical implications of his theory for the nature of sensations, or qualia. On this issue it does appear that an illusionist interpretation is available. Unlike Wimsatt, Globus maintains a dualism between representational consciousness and consciousness per se; they

are “two distinct classes” (298). More precisely, “that I am conscious is quite distinct from the content of consciousness in the sense that a context is distinct from that which is within the context or a constraint distinct from that constrained” (*ibid*). A similar dualism was apparent in Rosenberg’s version of panpsychism, discussed in chapters 2 and 3. He distinguished qualia, or “effective intrinsic carriers” of causal properties, from subjectivity or awareness itself, the “receptive carrier” of a context in which the state of a part of the universe is fully determinate. McKittrick complained that there seemed to be an ambiguity in Rosenberg’s system as to how exactly the indeterminate ‘background’ of possibility space prior to its “ingression” into a context of particular causal relations ought to be understood. Is it an event taking place in time, a transformation from objectively indeterminate physical conditions, or is possibility space more like a timeless conceptual necessity for providing a causal explanation? The former describes the most distant past; the latter describes the ultimate context of the present. Rosenberg was never entirely clear about which he meant,<sup>11</sup> which in turn makes it difficult to interpret his remarks about how the latest models of physical causation no longer require space-time; just what is it they are describing, and how does it mesh with phenomenal consciousness? Following a suggestion by Keith Turausky that the sum of all phenomenal properties might cancel each other out, Rosenberg put forward the idea that “level zero of reality” would consist of “a single experiential simple”. In chapter 3 I suggested that Rosenberg’s picture was highly reminiscent of Aristotelian theologians’ characterisation of ultimate reality as *nous*, in which matter cancels itself out at the highest level of conscious understanding (or, what amounts to the same thing, the highest context). However I went on to qualify this (in section [3.4]) with the observation that Rosenberg does *not* seek to identify ultimate reality with intelligibility, but rather with pre-reflective *feeling*,

---

<sup>11</sup> E.g. he describes the Turausky Fundamental Tone as a “carrier of effective properties” (Rosenberg 2017, p174), yet elsewhere denies that the background or “level zero” of the universe actually causes anything (Rosenberg 2015, p241).

and never explored the possibility that the “contentless openness” of pure receptivity, or consciousness, might also somehow coincide with the Turausky Fundamental Tone of pure *effectivity*, or phenomenal experience.

It seems to me that these dualisms can be collapsed into one another, and the resulting metaphysical model greatly simplified, if we accept an illusionist account of the genesis of qualia. After all, illusionists are not describing a temporal process in which intrinsic properties are literally *made* with the help of self-scanning mechanisms. There never were any such properties, only the illusion that there were, and that illusion can be explained by appeal to psychological facts about how we think, rather than an ontological account, such as those explored in chapter 2. Qualia would therefore be “confused thoughts”, which was how Descartes described sensations. Similarly, Rosenberg’s, Lockwood’s and Globus’s comments on how the non spatio-temporal nature of things themselves collapses into ‘Euclidian’ space-time upon observation should not be read as describing how things themselves *become* spatio-temporal. Otherwise the ‘pure’ consciousness of the meditator could not be said to embody things in themselves any better than the ordinary impure variety. For consciousness of nothing in particular, i.e. prime matter, to coincide with fundamental reality, consciousness of something must be a *type* of nothingness, a modification of it that only yields the illusion of richness and differentiation.

Echoing Globus and Rosenberg’s appeals to mystical states, Frankish writes that “it may be possible to dispel the illusion partially through indirect means, such as meditation and hypnotic suggestion; see, for example, Blackmore, 2011” (Frankish 2016, p18-19). Susan Blackmore’s 2011 book *Zen and the Art of Consciousness* is indeed an attempt to shore up thoroughgoing eliminative materialism with mystical insights, but I think her antipathy towards building any sort of consciousness into fundamental reality demonstrates a lack of imagination. At best, Blackmore shows that we can attain a point of total scepticism towards

*phenomenal* experience, and that what remains is neither conscious *nor* unconscious (Blackmore does, after all, report driving a car in such a state, p134). Like Blackmore, Frankish is also too quick to deny that phenomenal concepts can have any purchase on reality, writing that:

The concept of a mere introspectable *something*, which might or might not be qualitative, is not a genuine phenomenal concept, and if we conceptualized the properties of experience in that way, we would not feel any resistance to thinking of them as physical (a bare something might as easily be physical as nonphysical). (Frankish 2016, p25-6)

Frankish's point is conditional: *if* we experienced consciousness as just a 'something', then we could believe physicalism. Globus and others think this is achievable. But it couldn't be achievable if there were no corresponding phenomenal concept for such a state. Moreover, it is arguably a conceptual requirement for *something* to be directly present in order to ground the illusion of phenomenality, about which we enjoy Cartesian certainty. And to repeat, PCS has already provided us with a way of thinking about pure protoconsciousness; 'this' or 'that' were supposed to be our way of conceiving physical properties directly, as well as accommodating the inalienable sort of privacy that conscious experience was thought to have. And these 'thin' concepts can indeed be thought to constitutively incorporate the properties they represent. The only truly mind-independent property is best thought of as an indeterminate something about which nothing specific can be said, except that it somehow inspires us to make false judgements about the existence of phenomenal experiences and Euclidian objects.

Schroer also fits in with the form of neutral monism defended here. Far from avoiding an analysis of consciousness, Schroer's strategy actually gives one by distinguishing

phenomenal certainty from the revelation of any particular content or properties. His concept of translucent consciousness shows how ordinary, seemingly rich (“beefy”) consciousness could be constructed out of a combination of structural/relational properties which account for qualitative similarity relations, and intrinsic, but protophenomenal, properties for which we have no thicker concepts than bare demonstratives, such as saturation or hue in themselves. These latter opaque 'screen' elements of consciousness do not actually add up to full-blown consciousness; nor do they realize relational properties, because if they were conceived transparently we would realise that they are relational themselves, and the difference between aspects of experience that can be thought with phenomenal concepts and those that cannot would be undermined. This solves the palette problem by showing how the structuring of micro or protophenomenal properties with no clear and distinct phenomenality of their own gives rise to rich experience.

It also helps to solve a paradox in Raffman’s characterisation of qualia. Since there is no question of Raffman actually accounting for classic qualia, her insistence that there is more to them than can be captured by bare demonstratives is a sham; as Frankish would argue, the best a physicalist such as Raffman can do is account for zero qualia, which in turn are captured by thin demonstratives. In response, we can grant that Raffman qualia are, as she puts it, cognitively impenetrable if they are construed as classic. But at a higher level of description their properties are exhaustively conceptualised by the concept of zero qualia, which is indifferent to which exact shades it picks out. To see this, consider what would happen if colour nuances *were* penetrable: since the impossibility of doing so via phenomenal concepts gave rise to the explanatory gap, full cognitive penetration would dispel the illusion of intrinsic, physically inexplicable phenomenal content. We would be left with only physical properties, which, being nonphenomenal, could only be thought of as complex arrangements of ‘thises’, just as Schroer suggests. Rather than confirming that



mental qualities are mind-independent, as Heil and others argued, the existence of Raffman qualia presupposes illusionism about qualities. 'Raw' introspection would concern only those thin properties which it seems we can imagine zombies getting to know just as well as we can (there is no difference between mine and my zombie twin's use of 'that experience', since we are both being construed as equally blind as to what *that* is – which is precisely the reason Chalmers argues that this misconstrues the true nature of phenomenal concepts). Facts of the matter as to the presence or absence of classic qualia are then constructed by way of 'sharpenings', as Papineau terms them, carried out by the heuristics underlying introspection. Pre-reflectively, our phenomenal concepts were demonstrative of zero qualia all along, so the fact that Raffman has shown there can be no 'beefy' concepts of Raffman qualia is neither here nor there. This was already potentially the upshot of distinguishing experience from consciousness, as argued in chapter 2, and of making aspects of the mind "noumenal" in the course of Goff's attempts to avoid constitutive panpsychism, but by explicitly seeking to avoid any mention of qualia, the position I am recommending avoids making ontological commitments about the nature of phenomenal bonding. Experience could be one, many, both, or neither, and we would not notice the difference, so long as experience has no fundamental properties except for pure presence.<sup>12</sup>

Pure consciousness cannot be understood as a sophisticated form of subjective self-representation; the universe as we now understand it is much stupider than that (this point relates to how the medieval picture of the cosmos as the mind of God came to be usurped by a concept of autonomous, but unthinking, matter, which Descartes sought to defend). It would be better thought of as a pre-reflective state that could just as well be unconscious. Yet the recovery of this basic form of protoexperience might still be effected by the discovery of the *cogito*, which is not a form of self-representation, but a precondition for it.

---

<sup>12</sup> That presence need not be construed as presence to a subject in any metaphysically weighty sense, and so panpsychists can countenance Lichtenberg's analysis of experience, as discussed in [1.2.2].

As such, exercising reflective doubts about the nature of what appears to us, although a highly sophisticated form of self-consciousness, could still act as a methodological corrective to common sense in order to recover our original, pre-reflective condition of ignorance, blind habit, and featureless presence.

### **[6.5] Some Epistemological Problems**

Hume travels from scepticism to faith in appearances, while Descartes grounds science and philosophy in rational ideas. What, then, ought to be the epistemological status of appearances according to neutral monism? The rationalist solution to modal dualism says that we understand what is possible by seeking explanations that go beyond mere appearances, and situate them in the context of a theory. Hume (following Pyrrho, according to Fosl) blocks this by limiting theory to the description of constant conjunctions among appearances, rather than providing explanations for why things must be a certain way. The solution is to doubt what appearances are, and to question if we have a firm grip on where theory ends and appearances begin: theorisation would then be a (descriptively) inevitable aspect of our psychological relations to things in themselves. With regards to phenomenal properties, this would mean we cannot grant them the autonomy they seem to have. But we then have a problem of explaining what else could ground theory, if not sensory qualities, which according to Lockwood were all we could be sure of (see ch.2, n.10, above).

First, to recap, I submit that *sceptical doubt* is how we 'become' the world, in the sense meant by Aristotle, in the context of living in a reality that lacks intrinsically intentional properties, a universe that could not be thought of as knowing anything before life came along. This is a development of the dialectical solution to Aristotle's own antinomies, as

discussed in chapter 3, whereby ordinary experience can be recognised as, paradoxically, representing reality exactly as it is when its finitude, and hence apparent separation from the eternal form of the unmoved mover, is understood as the only way in which the unmoved mover could appear to itself. This solution is 'dialectical', in a sense later associated with Hegel, not only by virtue of seeking a solution to various dualisms, but by its tolerance of paradox, its banishment of noumenal things in themselves, subsumption of particulars to universals, and the claim that substance is subject, the God's eye view (Pinkard 2002, p253-4). For Moevs, dialectical resolution of paradox came from the attempt to purify the mind's mirroring of reality by the overcoming of the senses and "proximate matter" (Moevs 2005, p64). Naturally if this obstacle could not be overcome (the mind's mirror never adequately polished) by virtue of the nature of reality itself, then intellectual intuition could only be achieved through increasing one's proximity to proximate matter's unintelligibility. The alternative, radically materialist solution goes in reverse to the Medievals, where, paradoxically, reality is directly encountered when it is known *as* unknowable.

Thus, whereas Descartes discovers what one is, Hume tells us what we are doing. As Hume taught us, we discover at the end of doubt that the mind is passively and indifferently subject to the forces of habit which, through the association of ideas, lead us inexorably back to common sense. It has been said that reason's inability to rise to the challenge posed by its constitutive misrepresentation of experience as a source of justification for our beliefs demonstrates its weakness. Our failure to be terrified by our radical error – the error of believing beliefs are ever justified rather than merely caused - is itself part of the "nightmare" that Hume's scepticism leaves us with according to Wayne Waxman (1994, p267). Descartes gave us a more optimistic picture, insisting that reason can penetrate past the veil of confused ideas after all. Whereas Hume denied both selfhood and rationality, Descartes founded his worldview on them. But these are two halves of the same coin. As I

noted in [5.4], it looks as if whatever experience and intentionality are, they will ultimately be the sorts of things that can be studied from a third-person perspective, and their identity with natural facts will just be trivial. Yet there seems to be room for doubts about whether such mental entities exist, and about whether they could ever be naturalised.<sup>13</sup>

The perspective I am recommending credits both intuitions. For those still attached to the idea that intentionality requires teleology, and that the empirical world is roughly identical to our experience of it, the triumph of Cartesian metaphysics and mechanism will be equivalent to the metaphysical truth of scepticism. But for those already largely convinced that something like teleosemantics must be true, the discovery of a minimal self that is neither thought nor feeling provides the root from which mechanistic cognition could conceivably grow. I will tackle both ways if looking at things in turn.

### **[6.5.1] Scepticism**

It may seem as if there is a contradiction at the very outset of seeking to naturalise the mind, for it will inevitably reduce norms of truth and correctness, by which to distinguish genuine representation from misrepresentation, to mere descriptions (Kim 1988). Armed with just these descriptions we will have no reason to identify one set of mental states as true, from either a 1<sup>st</sup> or 3<sup>rd</sup> person perspective, and no reason to prefer such states to others – though, descriptively, we may have non-normative preferences regardless. And given that this seems to be the *upshot* of naturalism, there is no way for it to even get started, since we will have no basis for which to prefer a naturalistic description of mind to any other kind. This will be particularly vivid if such a description rules out our ever being in direct contact with the

---

<sup>13</sup> This dilemma also confronted Goff's panpsychism, where mind was supposed to be conceivable independently of any substance (as per the negative half of his project, against constitutive panpsychism and physicalism), and yet identical with the very stuff of substance itself, as per the metaphysics of Russellian monism.

world, for then we will never know if anything corresponds to that description, never mind whether we ought to believe it on the basis of those facts (“always believe theories based on robust evidence” is of course a norm, not something we can read off the facts). As Barry Stroud (2002) has pointed out, there is a school of epistemology that is, at least, immune to the objection from indirect realism. That is semantic externalism, which denies that we must know that we know anything, in order to know anything (Stroud, 2000, p109-113). And it is the latter which grounds naturalists’ claims to be realists about intentionality, against the objections of their critics. Since for externalists knowledge is so external that we don't have it anymore, it comes to nothing whether it is construed as a sceptical position, and likewise, whether Hume’s position is construed as an externalist one, since we could neither know nor feel the difference.

Stroud associates the constraint of knowing that you know with “traditional” epistemology (*ibid*, 99-101), which is supposed to be presuppositionless – since any presuppositions could be challenged by the sceptic (*ibid*, p104). By contrast, externalism “would explain knowledge in terms of conditions that are available from an ‘external’, third-person point of view, independent of what the knower's own attitude towards the fulfilment of those conditions might be” (*ibid*, p111). Stroud goes on to argue that the externalist criteria, though perhaps preferable to Descartes’ appeal to God, leave each of us stranded: what we know, what knowledge is, is always somebody else’s problem. For his part, then, Stroud is ambivalent: Suppose we denied that Descartes needed clear and distinct ideas to have knowledge, or that he needed a clear and distinct idea of God for God to guarantee that Descartes’ efforts to attain knowledge were not in vain. Then, Stroud says, the circularity Descartes would be left with, in appealing to God to make true all his beliefs, including the belief that God exists, would be unproblematic. But he also insists that that seems to be an unacceptable outcome.

Of course any explanation for why we know anything without knowing that we do will suffice here, so the naturalist can just substitute the relevant belief-forming mechanisms for God. Indeed, our predicament of being “externalist Cartesians” looks to be inevitable once the project of naturalising mental representation has so much as gotten off the ground. As Van Gulick (1980/1990) points out, birds' warning calls to each other constitute an implicit recognition of each other as capable of acquiring information about nearby predators, without their knowing this themselves (p124). He elaborates the point with regards to the extended mind thesis – a community, from bees to scientists, might embody knowledge that no individual possesses (*ibid*, p125-6). With regards to the processes underlying our own representations – states which we at least know we are in – nothing could be further from our awareness than *how* perception's causal relationship with the world is established: “A major reason experimentation is required to understand mechanisms is that most mechanisms, whether naturally occurring or manufactured by humans, do not reveal their parts, operations, or organization to the observer or user” (Bechtel 2007, p37).

The difficulty comes from seeking an account of what knowledge is. If how science knows is not the way traditional epistemologists thought knowledge should work (as self-reflective endorsement), why should we assume a description of cognition would be the exception that proves the old tradition was right? Stating a theory of what theories in general are should be the one thing science can't be expected to do. Yet even Stroud's imagined 'externalist Descartes', despite having abandoned the pretence to know what he does or does not know, “is, in addition, a theorist of knowledge” (Stroud 2002, p115). Stroud (116) objects, not due to his own allegiance to criteria of knowledge as knowing that you know, but due to the absence of any justification – any criteria, that is – for the externalist Cartesian to endorse his own claims.

I am aware that describing what I see as the deficiency in this way is not really satisfactory or conclusive. It encourages the 'externalist' to re-apply his theory of knowing or having good reason to believe at the next level up, and to claim that he can indeed understand himself to have good reason to believe his theory because he has good reason to believe that he does have good reason to believe his theory. (*ibid*, p118)

Stroud suggests (121) that the traditional epistemological project is doomed to dissatisfaction. Perhaps externalism is the best we can do. That leaves us with no reason to prefer mechanistic explanations over other kinds. However, if Gunderson is right, we can at least know a priori that the project of representing representation is impossible. That gets us some way towards accepting the naturalistic picture which predicts, on empirical grounds, that we should be in that predicament.

Moreover, the idea that knowledge consists in exercising an unconscious competence, if it came to be accepted, would amount to experiencing ourselves as being pushed around by unknown forces. That experience can also be arrived at by a priori considerations, thanks to Hume. According to Wayne Waxman (1994),

Hume's examination of identity, with its many references to the infallibility of immediate consciousness, seems to lead to the conclusion not merely that our belief in body *may* be false, but that it most *certainly* is. For, if it is inconceivable "that our senses shou'd be more capable of deceiving us in the situation and relations, than in the nature of our impressions" ([Hume 1978] 190), it follows that we not only *can* but *do* know - and know *infallibly* - that the objects immediately present to consciousness are *not* the continued, distinct existents our imaginations represent them to be. (Waxman 1994, p267-8)

Waxman's position is somewhat complicated by the fact that the supposed infallibility of immediate consciousness is really a species of psychological compulsion to assent (*ibid*), and by the wide gulf that separates Hume's definition of sense-experience from qualia (*ibid*, p58-62); sense-experience being more akin to proto-conscious 'experience' (as discussed in [2.2.1]). But these points will already have been conceded if we agree with a heuristic or mechanistic account of introspection. A more interesting point of overlap is between Hume and Descartes, since both philosophers exercised reflective scepticism towards the feelings of vivacity associated with the objects apparently disclosed to us via experience.

the less we allow such feelings to muddle our apprehension of the reality actually before us, the truer our picture of it will be. [Hume's] scepticism therefore has exactly the same foundation as his naturalistic analysis of human understanding: a pre-imaginative, privileged viewpoint which, by disclosing what it is we *really* are aware of (the given, the successive data, "immediately present to us by consciousness"), permits us to investigate the psychology underpinning natural beliefs (i.e. the causal efficacy and the external objects to which we *believe* ourselves witnesses). The Cartesian nightmare develops directly out of, and is the inevitable consequence of, a naturalism predicated on the ability to poke through the curtain of natural belief and descry, with eyes unblinkered by natural sentiment, the actuality there before us (*ibid*, p274-5)

According to Waxman, the aim of this unsentimental process of self-reflection is to disclose what is variously termed pre-imaginative, pre-phenomenological, pre-representative, pre-experiential and pre-identity consciousness (*ibid*, p275-6); experience that is "not so much *before* our eyes as they are *our very eyes themselves* - the *medium* in which the objects of our attentive, directed gaze present themselves to us, the *materials* of which their appearance is composed" (*ibid*, p60). Moreover, "the actions of the mind no more occupy



time than they do space. There can be no other way to discover them other than analytically and experimentally (introspection is, for all intents and purposes, useless)” (*ibid*, p61). It is Hume’s sceptical analysis of imagination, as that which constitutively misrepresents sensation, and not his appeal to a sensory given, which grounds his philosophy. But the nature of the given itself falls out of this.

As is well known, Hume went on to deny that such discipline of sceptical reflection could be maintained consistently, or for long, concluding that “the constitution of the human mind is such that we cannot be reasonable” (*ibid*, n.3, p334-5). However, given the inconsistency with which Waxman makes Hume’s case against reason – he rejects “negative dogmatism” (268) as a form of foundationalism, yet “the fact that we can *know* that [our] reason is a false reason means that his anti-foundationalism has a foundation” (336) – I submit that there is a strong motive to read naturalist metaphysics back into what Hume is saying. Individuals may not consistently endorse the naturalist picture of the mind, but, as Ladyman and Ross (2007, p29; 209) emphasise, if we have already countenanced an externalist teleosemantic account of representations then there is no obstacle to conceiving of the scientific community itself as succeeding where individuals fail. And, as Montero emphasised, the extended mind of science does seem to support negative dogmatism about common sense reality: time, space, objects and even causation<sup>14</sup> seem to have fallen by the wayside. With the sanction of cutting-edge science, then, Hume’s scepticism can be viewed retrospectively as precisely the insight into reality that Globus took pure consciousness for.

How is Descartes to be reconciled with Hume? What happened to the ‘common sense’ upshot of Humean scepticism, which was the non-dogmatic *acceptance* of appearances as the only possible motivators for passively forming beliefs and habits? Fosl

---

<sup>14</sup> Norton (2003) argues the identification of any one factor as a cause, even in a probabilistic sense, misrepresents scientific explanation. He emphasises that Salmon’s account of causation is compatible with this so long as it is construed as descriptive rather than normative (p5-6), and as I emphasised in chapter 2, Salmon’s account is supposed to accommodate Hume’s attack on ‘hidden powers’. I say more about this below.

(2011) argued that there is no contradiction between nature and convention, as the latter is a “cultivation” of the former, and so can accommodate scientific revisions to common sense. Yet, by being a contingent way of conducting ourselves, the project of understanding the world via science and philosophy is liable to be reified into something separate from mere habit, and to cause us to forget that our relations with things are essentially causal rather than intentional. Previously I argued that Wimsatt’s position was too radical to be construed as straightforwardly continuous with common sense. The same can be said for Hume, for whom common sense was entirely non-rational; common sense came to the rescue against total scepticism in a form that few people would recognise as either sensible or common. Both philosophers argue, from empirical and a priori grounds, respectively, that thought is not what we understood it to be.

This point extends to teleosemantics’ relationship with truth. Papineau’s teleosemantics leans heavily on the “redundancy theory of truth” (Papineau 1993, p83), whereby truth is to be explained away as a byword for conditions which satisfy our desires.<sup>15</sup> Anybody seeking more to truth than that will doubtless be disappointed, and in this respect Papineau’s account remains compatible with scepticism. But by putting his denial of some stronger form of truth on a metaphysical footing, it is possible to clear up some problems left outstanding, such as: why can’t we value truth in its own right, irrespective of its usefulness? As noted in chapter 3, Papineau (74-5) asserts that truth could be pursued in order to satisfy any desires, even “practically insignificant ones”. But his answer is unpersuasive given that, as a species of externalism, teleosemantics is doomed to leave us dissatisfied about how it is we know anything. In ruling out that higher-order perspective, it very much seems to deny, rather than accommodate, the pursuit of truth for its own sake, which motivated traditional epistemology. By seeing that pursuit as necessarily ending in

---

<sup>15</sup> Similarly, Salmon’s account is supposed to be a redundancy theory of causation, whereby ‘cause’ ceases to be a primitive term – hence his compatibility with teleofunctionalism.

disillusionment, and moreover disillusionment whereupon the self comes to seem amenable to naturalistic reduction, the gap is closed, and it becomes clear that, if there is any truth to be had, it can only be identified as whatever satisfies our desires.

The other issue this approach solves is that of “accidental replicas” (Papineau 1993, 90-94), beings atom-for-atom identical with humans, conjured out of thin air, which nevertheless lack intentional relations with their environment. As Papineau emphasises, a teleological account of representation leans heavily on the power of evolution to set trends as to what qualifies as a satisfied desire. Without prior precedent, an otherwise familiar physical state (a person commenting on the weather) will not ‘be’ any state in particular, and so that subject’s words and activities will have no meaning. Papineau treats these as *exotica* wholly separate from our own epistemic situation – beings we can barely imagine. However, we can perfectly well consider ourselves independently of our causal history and can at least fleetingly attain a state of extreme scepticism; the Pyrrhonian sceptic possesses the same propositional attitudes as his accidental replica. From a certain point of view, then, we are like the replicas, indeed, such a state could be our ‘original position’ prior to our acquiring the first associations of ideas at some point in early life (Waxman p275-6 identifies pre-conceptual consciousness with the “seedling” out of which the organism springs). Rather than scepticism putting us in contact with the sense data from which all beliefs are derived – associationism which is now regarded as implausible – the pre-conceptual is better off understood along the lines Globus suggests, as a state prior to any context whatsoever. Habit-formation starts prior to the world as we experience it, in the operations of the nervous system. In line with externalism, then, 1<sup>st</sup> person reason can know the *source* of pre-conscious empiricism, but not the specifics (which must be left to others, as Stroud puts it).

The project of naturalising representations therefore need not be worried by objections, such as those from Kusch (2006). Kusch argues that biological normativity is a

metaphor, a matter of projecting our high valuation of staying alive, onto the functions of the body (Kusch 2006, p73-4). Naturalists could accept Kusch's critique, which aims to defend scepticism about meaning, so long as it is understood as taking place at the higher-order level of analysis of traditional epistemology.<sup>16</sup>

### **[6.5.2] Knowledge**

How does this line up with the argument, advanced at the start of this chapter, that symbolic imagination is the best way of theorising about things in themselves, if the only direct apprehension of them is neither a thought nor a phenomenal experience, but rather a state of 'pure' (proto)consciousness? Scepticism and physicalism are compatible, I argued, because while the sceptic seeks to bracket the idea that the world is actually intelligible, in the scientific image of thought as a brain process, the reach of thought itself is always already bracketed by our limitations in space and time. Perfect modelling of reality is metabolically impossible, hence our reliance on heuristics, which can only ever establish a finite amount of intelligibility. Having no clear idea of what lies beyond our best theories is just a permanent fixture of our condition. This gives independent, empirical grounds for considering the ultimate context of our existence as unintelligible to reason.

---

<sup>16</sup> Kusch later argues against Heil's attempt to naturalise intentionality by analysing it as a form of disposition. Kusch is replying to Martin and Heil (1998), arguing, firstly, that mistakes in reasoning can be rationalised as dispositions with errant (quus-style) rules, and hence, secondly, that the kinds of disposition partners (though infinite in number – this is Martin and Heil's response to Kripke's 'Humean' objection that dispositions would need infinite time to demonstrate their consistency) with which our reason manifests itself cannot be known in advance. Kusch writes that "The normativity objection responds to that challenge by saying that the pairs <156, 2> and <2, 156> are linked to the number 158 by means of a disposition line. Unfortunately, this answer is hopelessly metaphorical. And at no point are we told how we can cash in this metaphor" (123). Following on from chapter 3's point of contention that Heil is incompatible with indirect realism, I would add that meaning scepticism and inscrutability of what it is dispositions aim for are mutually supporting. So externalism helps support Kusch against Heil, here.

Let us suppose that we are closest to reality in moments of extreme doubt. This still leaves the question, *if* one wishes to form beliefs about reality, for reasons that no doubt are ultimately habitual, then how best to go about it? It is in this conditional sense that the Cartesian strictures of sceptical reflection, objectifying representation, and bracketing of qualitative properties recommend themselves as the most able to satisfy our desires. That mode of reflection gives us the most sophisticated scientific understanding of what is going on when we represent the world, wherein mental representation is itself understood as a kind of proto-science, involving the formulation of heuristic strategies or theories for interfacing with the world, and the manipulation of formal symbols, or computation. And finally, the mind arrived at in moments of pure scepticism does not yield properties which could not be represented in such austere terms. Against those naturalists who deny the power of logical conceivability altogether, I suggest a compromise, between rational reflection and symbolic reasoning as our access to the *form* of mind-independent reality, and heuristics, robustness and non-logical modality as the best methods of exploring the specific contents of our local patch of reality. In the former case reality is unresistant to austere reasoning; in the latter case it exceeds all our powers of imagination.

As I argued above, Wimsatt's reduction of reason to heuristics is too totalising to give any content to the possibility of mechanistic explanation and its' being more right than alternative 'ways of knowing'. Physicalists have a strong motive for denying there is more to the physical than structure/dynamics, and analytic metaphysicians have good reasons for preserving intrinsic properties and modal monism. The only reason why the negative conceivability of Cartesian scepticism should appeal to sensory imagination is that, as Mary the colour scientist shows, symbolic/theoretical imagination cannot even countenance phenomenal properties as a possibility. But if the presence of consciousness, the property

that physicalism could never predict, were just intrinsicity *as such*, like the thought that thinks itself, then there is no problem.

## **[6.6] Conclusion**

If all of the above is correct, there is no need for philosophy of mind to appeal to either phenomenal properties or a well-defined subject in order to understand consciousness. All there is is an impersonal 'something', shaped into a seemingly determinate Cartesian theatre containing rich phenomenal properties by the cognitive impenetrability of some of its elements. This something is neither experiential nor conceptual, but something more fundamental; yet it may be best conceived, if we wish to conceive of it, via theoretical imagination. Nevertheless, it is not entirely alien to inner life, and we may still be said to know it better from the inside than through the eyes of science. This is a properly neutral monism, and likely to satisfy neither physicalists nor their opponents. Nevertheless, I think it poses a challenge to panpsychists who believed they could construe consciousness as a property of mind-independent reality without doing violence to it. Moreover, the emptiness of zero qualia avoids having to solve the bonding problem, since 'this' has no determinate boundaries or content that could clash with my ambiguous, functional unity – the only kind biology allows for.

Criticism of this view will have to come from outside the framework in which panpsychism operates. The power of Russell's solution to the mind-body problem was originally that there is some third concept that is neither physical nor phenomenal, but which grounds both. But the pursuit of such a rarefied type of stuff is essentially a reductionist line of enquiry, and so motivates illusionism about consciousness as much as

about the Euclidian properties of space and time. We should also bear in mind the comments by Russell about the implications neutral monism has for existential philosophy, namely, that “Although metaphysical materialism cannot be considered true, yet emotionally the world is pretty much the same as it would be if the materialists were in the right” (Russell, 2006, p162-3). Russell’s comments seem to undermine the idea that there are actual stakes to the mind-body problem outside of metaphysics, if the only options are panpsychism and physicalism; the dispute seems to be purely scholastic. But it seems to me that there are stakes, and that neutral monism simply does not address them. Goff raises the possibility in his book (2017a, p10) that having greater awareness that consciousness is an independent source of epistemic authority from science will free us from seeking technological solutions to ethical problems. But it is difficult to see how panpsychism can shore up that kind of epistemic authority, or how Goff’s cosmopsychism could re-enchant nature – he certainly sought to downplay any religious connotations it might otherwise seem to have. Nevertheless, it may be that whatever alternative conception of consciousness would come to the rescue at this point might be better off discarding the concept of ‘qualia’ anyway, at least if they are construed along the lines of sense-data, sensations, or “mental paint” (Block 2003). But that is a line of enquiry for another time.





## REFERENCES

- Abram, D. (1997) *The Spell of the Sensuous: Perception and Language in a More-Than-Human World*. New York: Vintage Books, Random House.
- Allison, H.E. (2004). *Kant's Transcendental Idealism: Revised and Enlarged Edition*. Yale University Press.
- Ameriks, K. (1990) Kant, Fichte, and Short Arguments to Idealism. *Archiv für die Geschichte der Philosophie*, 72, 65-85.
- Aristotle, Hardie, R.P. and Gaye, R.K. (trans.) (1984), *Physics*. In Barnes, J. (ed), *The Complete Works of Aristotle (Revised Oxford Translation): Volume One and Two*. Princeton University Press.
- Aristotle, Ross, W.D. (trans.) (1984), *Metaphysics*. In Barnes, J. (ed), *The Complete Works of Aristotle (Revised Oxford Translation): Volume One and Two*. Princeton University Press.
- Armstrong, D. (1961). *Perception and the Physical World*, London: Routledge & Kegan Paul.
- Austin, J.L.A. (1962) *Sense and Sensibilia*. Oxford University Press.
- Balog, K. (2011a) Acquaintance and the Mind-Body Problem. In: Hill, C. and Gozzano, S. (eds.), *New Perspectives on Type Identity*, Cambridge University Press, pp16-42.
- Balog, K. (2011b) Phenomenal Concepts. In: McLaughlin, B, Beckermann, A, Walter, S (eds.), *The Oxford Handbook of Philosophy of Mind*. Oxford University Press, pp292-313
- Balog, K. (2012) In Defense of the Phenomenal Concept Strategy. *Philosophy and Phenomenological Research*, Vol. LXXXIV No. 1, January, pp1-23
- Bechtel, W. (2008) *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. Routledge.
- Bett, R. (2011) How Ethical Can an Ancient Skeptic Be? In: Machuca, D.E. (ed.), *Pyrrhonism in Ancient, Modern, and Contemporary Philosophy*. Springer, pp3-18
- Blackmore, S. (2011) *Zen and the Art of Consciousness*. London: Oneworld Publications.
- Block, N. (1997) On a Confusion about a Function of Consciousness. In: Block, N, Flanagan, O, Güzelde, G (eds.), *The Nature of Consciousness: Philosophical Debates*. MIT Press, pp375-416
- Block, N. (2003) Mental Paint. In: Hahn, M. and Ramberg, B. (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Cambridge, Massachusetts: MIT Press, pp165-200
- Block, N. (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences* (30). Oxford University Press, pp481–548.
- Block, N. (2011) Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences* (20). Elsevier, 1-9.
- Boulter, S. (2007) *The Rediscovery of Common Sense Philosophy*. Palgrave Macmillan.

- Boulter, S. (2011) The Medieval Origins of Conceivability Arguments. *Metaphilosophy* (42), No. 5, October, 617-641.
- Boulter, S. (2013a) Aquinas on Biological Individuals. *Philosophia* (41), 603–616.
- Boulter, S. (2013b) *Metaphysics from a Biological Point of View*. Palgrave Macmillan.
- Boyle, R., Birch, T. (1672/1772) *The Works of the Honourable Robert Boyle*, vol. IV. London, Printed for J. and F. Rivington.
- Bradie, M. (1996) Ontic Realism and Scientific Explanation. *Philosophy of Science* (Proceedings), 63, 315-321.
- Burns, E.M., and Ward, W.D. (1982) Intervals, scales, and tuning. In: Deutsch, D (ed), *The Psychology of Music*. New York: Academic Press, pp 241–269.
- Burnyeat, M.F. (2003) Is an Aristotelian Philosophy of Mind Still Credible? (A Draft) In: Nussbaum, M.C., Rorty, A.O. (eds.), *Essays on Aristotle's De Anima*. Oxford University Press, pp15-26.
- Carnap, R. (1936). Testability and meaning. Part 1. *Philosophy of Science*, 3, 420–468.
- Carnap, R. (1937). Testability and meaning. Part 2. *Philosophy of Science* 4, 1–40.
- Carruthers, P. and Veillet, B. (2007) The Phenomenal Concept Strategy. *Journal of Consciousness Studies*, 14, no.9-10, 212-236.
- Chalmers, D. (1995/2010) Facing up to the Problem of Consciousness. *Journal of Consciousness Studies* (2), p200–19; reprinted with a new afterword in Chalmers, D., *The Character of Consciousness*, Oxford University Press, pp3-34.
- Chalmers, D. (1996) *The Conscious Mind: in search of a fundamental theory*. Oxford University Press.
- Chalmers, D. (2002a) Consciousness and its Place in Nature. In: Chalmers, D. (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford, New York: Oxford University Press, pp247–272.
- Chalmers, D. (2002b) Does conceivability entail possibility? In: Gendler, T. and Hawthorne J. (eds.), *Conceivability and Possibility*. Oxford: Oxford University Press, pp145–200.
- Chalmers, D. (2010a) Phenomenal Concepts and the Explanatory Gap. In Chalmers, D., *The Character of Consciousness*, Oxford University Press, pp305-336.
- Chalmers, D. (2010b) The Two-Dimensional Argument against Materialism. In: Chalmers, D., *The Character of Consciousness*, Oxford University Press, pp141-206.
- Chalmers, D. (2013) The Contents of Consciousness: Reply to Hellie, Peacocke, and Siegel. *Analysis* 73.2, 345–368.
- Chalmers, D. (2015) Panpsychism and Panprotopsychism. In: Alter, T., Nagasawa, Y. (eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford University Press, pp246-276.
- Chalmers, D. (2017) “The Combination Problem for Panpsychism” in Bruntrup, G & Jaskolla, L (eds.), *Panpsychism: Contemporary Perspectives*. Oxford University Press, pp179-215.

- Chisholm, R. (1957) Intentional Inexistence. In: Chisholm, R, *Perceiving: A Philosophical Study*, London: Oxford University Press, pp.168-185.
- Chomsky, N. (1988) *Language and Problems of Knowledge: The Managua Lectures*. Eleventh printing, 2001. London, England/Cambridge, Massachusetts: MIT Press.
- Churchland, P. M. (1981) Folk Psychology and the Propositional Attitudes. *Journal of Philosophy* (78), no. 2, 67-90.
- Churchland, P. M. (1985) Reduction, Qualia and the Direct Introspection of Brain States. *The Journal of Philosophy*, Vol. 82, No. 1 (Jan.), 8-28.
- Churchland, P. M. (1998) Knowing Qualia: A Reply to Jackson. In: Churchland, P. M., and Churchland, P. S., *On the Contrary*. MIT Press, pp143-158.
- Churchland, P. S., (1996) The Hornswoggle Problem. *Journal of Consciousness Studies* 3, 5-6. Exeter, Imprint Academic, 402-8.
- Cohen, S. (1978) Essentialism in Aristotle. *The Review of Metaphysics* Vol. 31, No. 3, Mar., 387-405.
- Cooper, W.E. (1990). William James's Theory of Mind. *Journal of the History of Philosophy*, Volume 28, Number 4, October, 571-593.
- Copi, I. (1954) Essence and Accident. *The Journal of Philosophy*, Vol. 51, No. 23, December 28-30, 706-719.
- Cottingham, J., Stoothoff, R., Murdoch, D., Kenny, A. (trans.), Descartes, R. (1984) *The Philosophical Writings of Descartes Vol. II*. Cambridge: Cambridge University Press.
- Cottingham, J., Stoothoff, R., Murdoch, D., Kenny, A. (trans.), Descartes, R. (1991) *The Philosophical Writings of Descartes Vol. III*. Cambridge: Cambridge University Press.
- Crane, T. (introduction), Dennett, D., Papineau, D. (2017) Papineau vs. Dennett: a philosophical dispute. *Times Literary Supplement*, August 4, no.5966.
- Cummins, R. (1975), 'Functional Analysis', *Journal of Philosophy*, 72: 741–65.
- Davidson, D. (1980) Mental Events. In: Davidson, D. *Actions and Events*. Oxford: Clarendon Press.
- Davies, B. (1992). *The Thought of Thomas Aquinas*. Oxford: Clarendon, Oxford University Press.
- Dennett, D.C. (1991) *Consciousness Explained*. Back Bay Books/Little, Brown and Company.
- Dennett, D.C. (2001) The Fantasy of First-Person Science. Unpublished; "a written version of a debate with David Chalmers, held at Northwestern University, Evanston, IL., February 15, 2001, supplemented by an email debate with Alvin Goldman".
- Descartes, R., Cottingham, J. and Williams, B. (trans.) *Mediations* [excerpts from Meditations II & IV] (1996), reprinted in Gertler & Shapiro (2007), *Arguing About the Mind*. London: Routledge, pp164-169
- Dowe, P. (1992) Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory. *Philosophy of Science* (59), 195-216.

- Esfeld, M. (2010) Humean metaphysics versus a metaphysics of powers. In: Ernst, G. and Hüttemann, A. (eds.), *Time, Chance and Reduction: Philosophical Aspects of Statistical Mechanics*. Cambridge University Press, pp119-135.
- Esfeld, M. (2017) The reality of relations: the case from quantum physics. In: Marmodoro, A. and Yates, D (eds.), *The Metaphysics of Relations*. Oxford: Oxford University Press, pp218-234.
- Feigl, H. (1958) The 'Mental' and the 'Physical'. In: Scriven, M., Maxwell, G. and Feigl, H. (eds.) *Minnesota Studies in the Philosophy of Science, Volume II: Concepts, theories, and the mind-body problem*. Minneapolis: University of Minnesota Press., pp320-492
- Feigl, H. (1960) Mind–body, not a pseudo-problem. In: Hook, S. (ed.) *Dimensions of Mind*. New York: New York University Press, pp24-34.
- Fosl, P.S., (2011) Skepticism and the Possibility of Nature. In: Machuca, D.E. (ed.), *Pyrrhonism in Ancient, Modern, and Contemporary Philosophy*. Springer, pp145-170.
- Frankish, K. (2012) Quineing Diet Qualia. *Consciousness and Cognition* (21), 667–676.
- Frankish, K. (2016) Illusionism as a Theory of Consciousness. In: Frankish, K., (ed.), *Illusionism*. Special Issue of *Journal of Consciousness Studies: controversies in science & the humanities*, Vol. 23, No. 11-12, November/December, pp11-39.
- Franklin, J. (2002) Stove's Discovery of the Worst Argument in the World. *Philosophy*, no. 77, pp. 615-24.
- Frege, G. (1948) On sense and reference. *Philosophical Review* (57), 209-30.
- Galavotti, M.C. (1999) Wesley Salmon on Explanation, Probability and Rationality. In: Galavotti, M.C., Pagnini, A. (eds.), *Experience, Reality, and Scientific Explanation: Essays in Honor of Merrilee and Wesley Salmon*. Springer Science & Business, pp39-54.
- Globus, G. (1972) Biological Foundations of the Psychoneural Identity Hypothesis. *Philosophy of Science* (no.39, September), 291-300.
- Globus, G. (1976) Mind, Structure, and Contradiction. In: Globus, G., Maxwell, G., and Savodnik, I. (eds.), *Consciousness and the Brain*. New York: Plenum, pp271-294.
- Goff, P. (2006) Experiences don't sum. *Journal of Consciousness Studies* (13).
- Goff, P. (2009a) Can the panpsychist get around the combination problem? In: Skrbina (ed.), *Mind that Abides: Panpsychism in the new Millennium*. John Benjamins B.V., pp129-135.
- Goff, P. (2009b) Why Panpsychism doesn't Help Us Explain Consciousness. *Dialectica* Vol. 63, N° 3, 289–311.
- Goff, P. (2010) Ghosts and Sparse Properties. *Philosophy and Phenomenological Research* (Vol. LXXXI No. 1, July), 119-139.
- Goff, P. (2012) A Priori Physicalism, Lonely Ghosts and Cartesian Doubt. *Consciousness and Cognition* (21), 742–746.
- Goff, P. (2013) Against Funny Physicalism. Unpublished draft, accessed on 04 October 2013, 17:57 at: [https://www.academia.edu/3827649/Against\\_funny\\_physicalism](https://www.academia.edu/3827649/Against_funny_physicalism)

- Goff, P. (2014a) Orthodox Property Dualism + The Linguistic Theory of Vagueness = Panpsychism. In: Brown, R. (ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*. Springer Science, Studies in Brain and Mind (6), pp75-91.
- Goff, P. (2014b) The Cartesian Argument against Physicalism. In: Sprevak, M. and Kallestrup, J. (eds.), *New Waves in Philosophy of Mind*. Palgrave Macmillan, pp3-20.
- Goff, P. (2014c) The Phenomenal Bonding Solution to the Combination Problem. Unpublished draft, accessed on 10 November 2014, 16:15 at:  
[http://www.philipgoffphilosophy.com/uploads/1/4/4/4/14443634/the\\_phenomenal\\_bonding\\_solution\\_to\\_the\\_combination\\_problem.pdf](http://www.philipgoffphilosophy.com/uploads/1/4/4/4/14443634/the_phenomenal_bonding_solution_to_the_combination_problem.pdf)
- Goff, P. (2014d) *Consciousness and Fundamental Reality*, chapter 7: An Elegant Theory of Matter. Unpublished draft, accessed on 17 May 2016, 18:05 at:  
[http://www.philipgoffphilosophy.com/uploads/1/4/4/4/14443634/chapter\\_7\\_-\\_an\\_elegant\\_theory\\_of\\_matter.pdf](http://www.philipgoffphilosophy.com/uploads/1/4/4/4/14443634/chapter_7_-_an_elegant_theory_of_matter.pdf)
- Goff, P. (2015a) Against Constitutive Russellian Monism. In: Alter, T., Nagasawa, Y. (eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*, Oxford University Press, pp370-400.
- Goff, P. (2015b) Real Acquaintance and Physicalism. In: Coates, P, and Coleman, S (eds.), *Phenomenal Qualities: Sense, Perception, and Consciousness*. Oxford University Press, pp121-145.
- Goff, P. (2017a) *Consciousness and Fundamental Reality*. New York: Oxford University Press.
- Goff, P. (2017b) The Phenomenal Bonding Solution to the Combination Problem. In: Brüntrup, G and Jaskolla, L (eds.), *Panpsychism: Contemporary Perspectives*. Oxford University Press, pp283-304.
- Goff, P., Papineau, D. (2013/4) What's wrong with strong necessities? *Philos Stud* (2014) 167/Springer Science+Business Media Dordrecht 2013, 749–762.
- Grahek, N., Dennett, D.C. (foreword), (2007) *Feeling Pain and Being in Pain*. MIT Press, 2<sup>nd</sup> Ed.
- Greene, B.R., (2000) *The Elegant Universe: superstrings, hidden dimensions, and the quest for the ultimate theory*. W. W. Norton & Company, Inc.
- Gunderson, K. (1970). Asymmetries and Mind-Body Perplexities. In: Radner, M. and Winokur, S. (eds.), *Minnesota Studies in the Philosophy of Science, vol. 4*, Minneapolis: University of Minnesota Press, pp. 273-309.
- Halsey, R.M., & Chapanis, A. (1951) Number of absolutely identifiable hues. *Journal of the Optical Society of America* (41), 1057-8.
- Hart, W.D. (1988/2007) An Argument for Dualism. Excerpt from Hart (1988), *Engines of the Soul*, Cambridge University Press; reprinted in Gertler, B., & Shapiro, L. (eds.) (2007), *Arguing About the Mind*. Routledge.
- Heil, J. (2003a). *From an Ontological Point of View*. Oxford: Clarendon.
- Heil, J. (2003b). Levels of Reality. *Ratio* (new series) XVI 3 September 2003 0034–0006, 205-221.
- Heil, J. (2004a). *Philosophy of Mind: A Contemporary Introduction*. Routledge.

- Heil, J. (2004b). Properties and Powers. In Zimmerman, D.W., *Oxford Studies in Metaphysics: Volume 1*. Oxford University Press, pp223-254.
- Heil, J. (2015). Aristotelian Supervenience. *Proceedings of the Aristotelian Society*, Vol. cxv, Part 1 (doi: 10.1111/j.1467-9264.2015.00383.x), 41-56.
- Heisenberg, W. (1962). *Physics and Philosophy: The Revolution in Modern Science*. New York: Harper & Row, Harper Torchbooks.
- Hempel, C.G. (1962). Deductive-nomological vs. statistical explanation. In Feigl, H. and Maxwell, G., *Minnesota studies in the philosophy of science, Vol. III: Scientific explanation, space, and time*. Minneapolis, MN: University of Minnesota Press.
- Hill, C. (1997) Imaginability, Conceivability, Possibility and the Mind-Body Problem. *Philosophical Studies* (87). Kluwer, 61–85.
- Hoefer, C. (2011) Time and Chance Properties. In: Callender, C. (ed.), *The Oxford Handbook of Philosophy of Time*. New York: Oxford University Press, pp68-90
- Horgan, T. (1984) Jackson on physical information and qualia. *Philosophical Quarterly* (34), 127-32.
- Horgan, T. (1993) From Supervenience to Superdupervenience: Meeting the Demands of a Material World. *Mind*, Vol. 102 (408, October), Oxford University Press, 555-586.
- Hume, D., Selby-Bigge, L.A., and Nidditch, P.H. (eds.), (1975) *David Hume's enquiries concerning human understanding and concerning the principles of morals*. 3rd edition. Oxford: The Clarendon Press.
- Hume, D., Selby-Bigge, L.A., and Nidditch, P.H. (eds.), (1978) *A Treatise on Human Nature*. Oxford: The Clarendon Press.
- Hurvich, L.M. (1981) *Color Vision*. Sunderland, MA: Sinauer Associates.
- Hutto, D. (2000) *Beyond Physicalism*. John Benjamins Publishing Co., Amsterdam/Philadelphia.
- Hutto, D., Myin, E. (2013) *Radicalizing Enactivism: Basic Minds Without Content*. MIT Press.
- Jackson, F. (1982) Epiphenomenal Qualia. *Philosophical Quarterly* (32), 127-36.
- James, W., Miller, G.A. (introduction), (1983) *The Principles of Psychology*. Harvard University Press.
- Jones, R.H. (2016) *Philosophy of Mysticism: Raids on the Ineffable*. State University of New York Press, Albany.
- Kim, J. (1988). What Is "Naturalized Epistemology?" *Philosophical Perspectives, Vol. 2, Epistemology*, 381-405.
- King, P. (1994). Scholasticism and the Philosophy of Mind: The Failure of Aristotelian Psychology. in Horowitz, T. And Janis, A.I., *Scientific Failure* (Rowman & Littlefield), pp109–138.
- Klima, G., Allhoff, F. and Vaidya, A. (2007). *Medieval Philosophy*. Malden, MA: Blackwell Pub.
- Kripke, S., (1980) *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.

- Kusch, M. (2006), *A Sceptical Guide to Meaning and Rules: Defending Kripke's Wittgenstein*. Acumen, p120-3
- Ladyman, J., Ross, D., Spurrett, D., Collier, J. (2007) *Every Thing Must Go: Metaphysics Naturalized*. Oxford University Press.
- Langland-Hassan, P. (2015) Introspective Misidentification. *Philosophical Studies*, vol.172, issue 7, 1737-1758.
- Langton, R. (2000) "The musical, the magical, and the mathematical soul", in Crane, T., and Patterson, S. (eds.), *History of the Mind-Body Problem*. London/New York: Routledge, p13-33.
- Levine, J. (1983). "Materialism and qualia: The explanatory gap". *Pacific Philosophical Quarterly*, 64(4), 354-361.
- Levine, J. (1993) On leaving out what it's like. In: Davies, M, and Humphreys, G. (eds.), *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
- Levine, J. (2001) *Purple Haze: The Puzzle of Consciousness*. Oxford University Press.
- Lewis, D. (1980) Mad pain and Martian pain. In Block, N. (ed.), *Readings in the Philosophy of Psychology*, vol. 1. Cambridge, MA: Harvard University Press.
- Lewis, D. (1981). *Philosophical Papers, Volume I*. Oxford: Oxford University Press.
- Lewis, D. (1983) New work for a theory of universals. *Australasian Journal of Philosophy*, 61/ 4, 343–77.
- Lewis, D. (1986) *Philosophical Papers, Volume II*. Oxford: Oxford University Press.
- Lewis, D. (1993). "Mathematics is Megethology", *Philosophia Mathematica* 1, 3–23. Reprinted in Lewis (1998), *Papers in Philosophical Logic* (Cambridge: Cambridge University Press), p203–29.
- Lewis, D. (2004) What Experience Teaches. In: Ludlow, P., Nagasawa, Y., and Stoljar, D. (eds.), Jackson, F. (foreword), *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*. MIT Press, pp77-104.
- Loar, B., (1990) "Phenomenal States", in Tomberlin, J (ed.), *Philosophical Perspectives*, Vol. 4. Atascadero: Ridgeview Publishing Co., pp81–108. Reprinted with alterations in Block, N, Flanagan, O, Güzeldere, G (eds.) (1997), *The Nature of Consciousness: Philosophical Debates*. MIT Press, pp597-616.
- Lockwood, M. (1990) *Mind, Brain, and the Quantum*. Basil Blackwell.
- Lockwood, M. (2015) The Grain Problem. In: Alter, T., and Nagasawa, Y. (eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford University Press, pp143-160.
- Maddy, P. (2007) *Second Philosophy: A Naturalistic Method*. Oxford University Press.
- Margenau, H. (1954) Advantages and Disadvantages of Various Interpretations of the Quantum Theory. *Physics Today* 7 (10) October, 6-13.
- Martin, C. B., Heil, J. (1998). "Rules and Powers". In *Language, Mind, and Ontology. Philosophical Perspectives* 12, pp283–312.

- Mayr, E. (1982) *The Growth of Biological Thought*. Cambridge, MA, Harvard University Press.
- McGinn, C. (1989) Can We Solve the Mind-Body Problem? *Mind*, Vo. xcvi, no. 391. Oxford University Press, July, 349-366.
- McKittrick, J. (2006) Rosenberg on Causation. *Psyche*, 12 (5), December, 1-21.
- McLaughlin, B. (1992) The Rise and Fall of British Emergentism. In: Beckermann, A, Flohr, H, Kim, J (eds.), *Emergence or Reduction? : essays on the prospects of nonreductive physicalism*. de Gruyter, pp49-93.
- Metzinger, T. (2003) *Being No-One: the self-model theory of subjectivity*, MIT Press.
- Mill, J. (1973) A System of Logic, Ratiocinative and Inductive. In Robson, J. (ed.), *The Collected Works of J. S. Mill* (Volumes 7–8). Toronto: University of Toronto Press.
- Millikan, R. (1984) *Language, Thought, and Other Biological Categories*. Cambridge, Mass.: MIT Press.
- Moevs, C. (2005) *The Metaphysics of Dante's Comedy*. Oxford University Press.
- Montero, B. (2001) Post-Physicalism. *Journal of Consciousness Studies* (8), No. 2, 61–80.
- Nagel, T. (1974) What is it like to be a bat? *Philosophical Review* 4:435-50.
- Nagel, T. (1979) Brain Bisection and the Unity of Consciousness. In Nagel, T., *Mortal Questions*. Cambridge, pp147-164.
- Nagel, T. (1986) *The View From Nowhere*. Oxford University Press.
- Nolan, D. (2005). *David Lewis (Philosophy Now)*. Routledge.
- Noonan, H. (2013). *Routledge Philosophy Guidebook to Kripke and Naming and Necessity*. Routledge.
- Norton, J.D. (2003) Causation as Folk Science. *Philosophers' Imprint*, Volume 3, No. 4 November, 1-22.
- Odenwald, S. (2002) *Patterns in the Void: Why Nothing is Important*. Basic Books.
- Pagels, H. R (1982). *The Cosmic Code: Quantum Physics as the Language of Nature*. New York: Bantam Books.
- Papineau, D. (1993) *Philosophical Naturalism*. Blackwell.
- Papineau, D. (2007a) Kripke's proof is Ad Hominem not Two-Dimensional. *Philosophical Perspectives*, 21, 475-494.
- Papineau, D. (2007b) The Case for Materialism. In: Gertler, G, Shapiro, S (eds.), *Arguing About the Mind*. Routledge, pp125-132.
- Parfit, D. (1971). "Personal Identity". *Philosophical Review*, 80: 3-27.
- Patterson, S. (2000) How Cartesian Was Descartes? In: Crane, T. & Patterson, S. (eds.), *History of the Mind-Body Problem*. Routledge, p70-110.



- Pinkard, T. (2002). *German Philosophy 1760-1860: The Legacy of Idealism*. Cambridge University Press.
- Popper, K. (1935/1959). *The Logic of Discovery*. London: Hutchinson.
- Putnam, H. (1975), The Meaning of 'Meaning'. In Putnam, H., *Philosophical Papers, Volume 2: Mind, Language and Reality*. Cambridge University Press.
- Putnam, H. (1981) *Reason, Truth and History*. Cambridge University Press.
- Putnam, H. (1995). Pragmatism. *Proceedings of the Aristotelian Society*, 95: 291–306.
- Pyysiäinen, I. (2006) Does Meditation Swamp Working Memory? *Behavioral and Brain Sciences* 29 (no. 6): 626–27.
- Quine, W.V.O. (1960). *Word and Object*. Cambridge, Mass.: MIT Press.
- Raffman, D. (1993) *Language, Music and Mind*. MIT Press.
- Raffman, D. (1995) On the Persistence of Phenomenology. In: Metzinger, T. (ed.), *Conscious Experience*. Imprint Academic; Schöningh, pp293-308.
- Rorty, R. (1979) *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.
- Rosenberg, G. (2015) Causality and the Combination Problem. In: Alter, T., Nagasawa, Y. (eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford University Press, pp224-245.
- Rosenberg, G. (2017) Land Ho? We Are Close to a Synoptic Understanding of Consciousness. In: Bruntrup, G & Jaskolla, L (eds.), *Panpsychism: Contemporary Perspectives*. Oxford University Press, pp153-176.
- Rowlands, M (2003) *Externalism*. Acumen.
- Russell, B. (1927) *The Analysis of Matter*. London: Kegan Paul.
- Russell, B. (1921) *The Analysis of Mind*. London: George Allen and Unwin.
- Russell, B. (1927/1992), *An Outline of Philosophy*. London: Routledge.
- Russell, B. (2006) *In Praise of Idleness and Other Essays*. Routledge.
- Salmon W. (1994) Causality without Counterfactuals. *Philosophy of Science* (61), pp.297-312.
- Salmon, W. (1977) A Third Dogma of Empiricism. In: Butts, R. And Hintikka, J. (eds.), *Basic Problems in Methodology and Linguistics*. Dordrecht: Reidel, pp. 149-166.
- Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

- Schilder, P., and Stengel, E. (1928). Schmerzsymbolie. *Zeitschrift für die gesamte Neurologie und Psychiatrie* 113:143–148.
- Schroer, R. (2010) Where's the Beef? Phenomenal Concepts as Both Demonstrative and Substantial. *Australasian Journal of Philosophy* Vol. 88, No. 3, 505–522.
- Schwitzgebel, E. (2014) The Crazyist Metaphysics of Mind. *Australasian Journal of Philosophy*, Vol. 92, No. 4, 665–682.
- Searle, J. R. (1980) Minds, Brains and Programs. *The Behavioral and Brain Sciences*, vol. 3. Cambridge, UK: Cambridge University Press, pp417-424.
- Searle, J.R., Dennett, D.C., Chalmers, D.J. (1997) *The Mystery of Consciousness*. New York Review of Books.
- Sellars, W. (1963) Philosophy and the Scientific Image of Man. In: Sellars, W., *Science, Perception, and Reality* London: Routledge.
- Shaviro, S. (2014) *The Universe of Things: On Speculative Realism*. University of Minnesota Press.
- Sider, T. (2012) *Writing the Book of the World*. Oxford University Press.
- Siegel, J.A. & Siegel, W. (1977a) Absolute identification of notes and intervals by musicians. *Perception & Psychophysics*, 21, 143-52.
- Siegel, J.A. & Siegel, W. (1977b) Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception and Psychophysics*, 2.1, 399-407.
- Simon, J. (2014) What is Acquaintance with Consciousness? In: Brown, R. (ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*. Springer Science, Studies in Brain and Mind 6, pp103-118.
- Skrbina D. (ed.) (2009) *Mind that Abides: Panpsychism in the new millennium*. John Benjamins Publishing Co., Amsterdam/Philadelphia.
- Smart, J.J.C. (1959) Sensations and Brain Processes. *Philosophical Review* (68), 141–156.
- Standish, R.K. (2006) *Theory of Nothing*. Booksurge Australia.
- Stoljar, D. (2005) Physicalism and Phenomenal Concepts. *Mind and Language*, (no. 20), 469–94.
- Stoljar, D. (2010). *Physicalism*. Hoboken: Taylor & Francis/ Routledge.
- Stone, M.W.F (2000). The soul's relation to the body: Thomas Aquinas, Siger of Brabant and the Parisian debate on monopsychism. In Crane, T., and Patterson, S. (eds.), *History of the Mind-Body Problem*. London/New York: Routledge, pp 34-69.
- Stove, D. (1991), Idealism: A Victorian Horror Story. In Stove, *The Plato Cult and Other Philosophical Follies*, Oxford, Blackwell, pp83-177.
- Strawson, G. (2004) Against Narrativity. *Ratio* 16 XVII 4 December, pp.428–52; reprinted in Strawson, G. (2008) *Real Materialism and other essays*, pp189-208.
- Strawson, G. (2008), *Real Materialism and other essays*. Oxford University Press.

- Strawson, G. (2009), Realistic Monism: Why Physicalism Entails Panpsychism. In: Skrbina, D (ed.), *Mind that Abides*. Advances in Consciousness Research vol. 75, John Benjamins, pp33-66.
- Stroud, B. (2002) Understanding Human Knowledge in General. In Stroud, B., *Understanding Human Knowledge: Philosophical Essays*. Oxford University Press.
- Tarnas, R. (2006) *Cosmos and Psyche: Intimations of a New World View*. Viking Penguin.
- Unger, P. (2009) I do not exist. In: Rea, M. (ed.), *Arguing About Metaphysics*. New York: Routledge, pp94–105.
- Valberg, J.J. (2007). *Dream, Death and the Self*. Princeton University Press.
- Van Gulick, R. (1980) Functionalism, Information and Content. *Nature and System* (2), 139–60. Reprinted in Lycan, W.G. (ed.) (1990) *Mind and Cognition: A Reader*. Cambridge, Mass. Basil Blackwell, 1<sup>st</sup> edition, pp107-129.
- Vinci, T., Robert, J.S. (2005). Aristotle and Modern Genetics. *Journal of the History of Ideas*, Vol. 66, No. 2 (April), 201-221.
- Waxman, W. (1994) *Hume's Theory of Consciousness*. Cambridge University Press.
- Wheeler, J. A. (1990) Information, physics, quantum: The search for links. In W. H. Zurek (ed.), *Complexity, Entropy, and the Physics of Information*. Redwood City, Calif.: Addison-Wesley, pp3-28.
- Whitehead, A.N. (1929, 1957, 1978), Ray Griffin, D. & Sherburne, D.W. (eds.), *Process and Reality: an essay in cosmology*. The Free Press, Macmillan Publishing.
- Wilkes, K.V. (1984) Is Consciousness Important? *Brit. J. Phil. Sci.* 35, 223-243.
- Wilkes, K.V. (1988) *Real People: personal identity without thought experiments*. Oxford University Press.
- Wilkes, K.V. (2003) *Psyche versus the mind*. In: Nussbaum, M.C., Rorty, A.O. (eds.), *Essays on Aristotle's De Anima*. Oxford University Press, p109 – 127.
- Wimsatt, W.C. (2007) *Re-Engineering Philosophy for Limited Beings: piecewise approximations to reality*. Harvard University Press.
- Wittgenstein, L., Anscombe, G.E.M. (trans.), (1953) *Philosophical Investigations*. Oxford: Basil Blackwell.
- Wittgenstein, L, McGuinness, B.F., Pears, D.F. (trans.), Russell, B., (introduction), (1974). *Tractatus Logico-Philosophicus*. Revised Edition; Routledge & Kegan Paul.
- Yovel, Y. (2018). *Kant's Philosophical Revolution: A Short Guide to the Critique of Pure Reason*. Princeton University Press.