2009

Consciousness From A Naturalistic Perspective

Hugh R. Alcock
Wilfrid Laurier University

CONSCIOUSNESS FROM A NATURALISTIC PERSPECTIVE

by

Hugh R. Alcock

MPhil, University of Birmingham, 2001
BA. Hons, University of Guelph, 1999

DISSERTATION
Submitted to the Department of Philosophy
in partial fulfilment of the requirements for
Doctor of Philosophy

Wilfrid Laurier University
April, 2009

Your file Votre référence
ISBN: 978-0-494-49965-8
Our file Notre référence
ISBN: 978-0-494-49965-8

ABSTRACT

Despite advances in neuroscience some sceptics, e.g., David Chalmers, Colin McGinn and Thomas Nagel, contend that we are no nearer to achieving a scientific understanding of phenomenal consciousness. These sceptics claim that naturalising consciousness, i.e., subsuming it under our scientific theories, is either impossible, at least without radically reforming our current scientific practices, or perhaps beyond our cognitive grasp. Their scepticism is based on what is called the 'problem of consciousness' or the 'hard problem'. I argue that their pessimism is unwarranted. Their conclusion is based on adopting a non-naturalistic attitude, according to which our scientific theories must accommodate our intuitive understanding of phenomena. Adopting this attitude, I argue, is a bad policy quite generally as it is liable to lead to unconstrainable metaphysical claims. Our best policy is to adopt a naturalistic attitude instead, characterised by thinking of philosophy as continuous with science, as W.V. Quine urged. And from this naturalistic perspective explaining consciousness in scientific terms is possible. Our phenomenological, i.e., first-personal, understanding of consciousness is based on qualia which, I argue, are unnaturalisable. However, there are two ways in which we understand consciousness, namely, naturalistically in terms of behaviour and physiology and phenomenologically, and I argue that although it may seem contradictory these ways of understanding consciousness are mutually dependent. Consequently, consciousness as it is understood naturalistically is bound up with our phenomenological understanding of it. Therefore, inasmuch as consciousness understood naturalistically is subsumable under our scientific theories so consciousness *simpliciter* is naturalisable.

ACKNOWLEDGEMENTS

Special thanks to my dissertation adviser Neil Campbell. I could not have hoped for a better adviser. Above all he has demonstrated to me the virtues of clarity and organisation in philosophical writing. Special thanks also to the other two members of my committee, Andrew Bailey and Rockney Jacobsen, who have been exemplary in their professionalism and generous with their time. Also I wish to thank my external examiner, Philip Dwyer (University of Saskatchewan), and my internal/external examiner, Peter Eglin (Department of Sociology). Both examiners had valuable insights into my work and have given me much to think about. In addition, I owe a sincere debt of thanks to Rhea and Ray Whitehead and Cathie Neagle, without whose support none of this would have been possible. Finally, many thanks to my parents Roy and Sheila Alcock for their support during my years of study.

CONSCIOUSNESS FROM A NATURALISTIC PERSPECTIVE

Hugh R. Alcock

TABLE OF CONTENTS

INTRODUCTION

As our understanding of the brain, and of cognition more generally, has advanced some have begun seriously to try to explain mental phenomena, that is, to offer a viable theory of mind. These efforts find support among many philosophers of mind; perhaps the most notable of whom are Daniel Dennett and Paul and Patricia Churchland. There could almost be said to exist a partnership between neuroscientists and philosophers in this respect.[1] At the same time, however, other philosophers see the possibility of a neuroscientific theory of mind as either extremely problematic or downright mistaken. This scepticism originates, most notably, with the work of Thomas Nagel (1974) and Frank Jackson (1982). Other more recent sceptics include David Chalmers (1996), Colin McGinn (1991), Galen Strawson (1994). Their concerns focus on phenomenal consciousness. They argue that our inability to think of consciousness in physical terms either demonstrates or strongly suggests that it is not a physical property and as such it is beyond the limits of science, at least as it is currently practised. In this dissertation I

[1] This partnership has led to the publication of many popular books expounding such theories, e.g., Antonio Damasio's *Descartes' Error* (1994), Daniel Dennett's *Consciousness Explained* (1991) and Nicholas Humphrey's *A History of the Mind* (1994), and more recently G.M. Edelman and G. Tononi's *A Universe of Consciousness* (2000) and Christof Koch's *The Quest for Consciousness* (2004).

consider the philosophical reasoning behind this scepticism about the possibility of naturalising phenomenal consciousness, i.e., capturing it in our scientific theories.

Phenomenal consciousness (henceforth 'consciousness') is that property constitutive of our phenomenal experiences, e.g., our seeing a blue sky, our smelling the fragance of a flower, our hearing the musical notes of a composition etc. In seeing a blue sky we do not simply see the sky as blue, but rather we *feel* it as such. Consciousness concerns how things feel or seem to us in this sense. Without consciousness, we suppose, perceiving the world would be mechanical; that is to say, there would be nothing more to seeing, hearing, touching things etc. than bodies reacting to these things in ways that could be wholly described in mechanical terms. And this is not how we think our experiences go. Our experiences are alive, they are conscious or felt. Pain, for example, is not merely a mechanical reaction to bodily damage, it also has the qualitative character of hurtfulness; besides involving a set of behavioural dispositions such as avoidance and exclamations, pain experiences have a raw feel to them.

We think of consciousness, therefore, as comprising the qualitative character of our phenomenal experiences. Each of the properties that characterise experience in this qualitative sense is often referred to as a 'quale'. Experiences differ to the extent that their qualia differ. For example, my experience of looking at a ripe red tomato differs from that of looking at an unripe green tomato at least to the extent that the former has a red quale while the latter has a green quale. Likewise, the experiences of a tingling and of a dull pain at least differ in having distinct qualia whatever behaviour is exhibited.

Only the person having some experience with these distinguishing properties can be acquainted with them. However hard we might observe another person we cannot be

acquainted with the qualitative character of their experiences as we suppose we can each be in our own case. I feel a toothache I am having, say, and no one else can do so. This fact about our experiences seems obvious enough. And because qualia are private in this way we understand them to be peculiar or special. All other properties in the world are publicly apprehensible. For example, the property of being a cat, call it 'catness', under certain conditions is apprehensible by me, you, and anyone else who grasps the concept. Most publicly accessible properties can be analysed in terms of other publicly accessible properties. For example, we can explain catness in terms of being furry, having four legs, and in other much more specialised terms concerning such things as having a certain type of DNA. Qualia, it would seem, are not open to analysis in this way.

This peculiarity of qualia, and consciousness more generally, points to the problem of consciousness. In various ways many have argued that if, as physicalism claims, everything is physical, then there should in principle be some way of understanding qualia, i.e., analysing them in terms of physical properties.[2] However, physical properties are quintessentially publicly accessible, that is, to be a physical property is to be apprehensible by anyone, either directly or indirectly; and so, if qualia cannot in principle be apprehended publicly, then they appear not to be reducible to physical properties. Qualia seem to be incommensurable with physical properties. This suggests, therefore, that qualia are not physical. Indeed, qualia are essentially thought of as perspectival in nature, that is, they exist for the subject rather than being something that the subject happens to apprehend. So, for example, a red quale can be described as the 'what it is like'

---

[2] Notable examples, discussed later, are Thomas Nagel (1974), Frank Jackson (1982), and David Chalmers (1996).

for the subject to see something as red. Qualia are the subjective aspects of our experiences in this sense. Moreover, this incommensurability between qualia and physical properties is suggested in another way.

As Chalmers observes, it is commonly supposed that consciousness is caused physically even though "we have no good explanation of why and how it so arises" (1995, 201). Or as he otherwise phrases the puzzle: "Why should physical processing give rise to a rich inner life at all?" (*ibid*). Here, by 'rich inner life' Chalmers is alluding to the peculiarly subjective nature of consciousness. His point is simple: qualia seem completely distinct from physical properties so that there is no clear way at all of relating them. Often Chalmers describes this incommensurability in terms of our always being able to conceive of consciousness independent of any physical properties. Thus, to the extent that conceivability is the same as logical possibility it is logically possible that qualia are distinct from physical properties. Assuming this is so, Chalmers argues, qualia are not identical with physical properties, i.e., they are not ontologically reducible to physical properties. That is because identity is a necessary relation, that is, if two things are identical then it is logically *impossible* for them to be distinct.[3]

This, the problem of consciousness, is usually delineated in terms of its being a problem for physicalism, especially since it leads to the denial that everything is ontologically reducible to physical properties. However, as adumbrated earlier, I want to focus on the problem as it concerns the possibility of explaining consciousness naturalistically. If consciousness is not reducible to physical properties, then it is not

---

[3] Here Chalmers follows Saul Kripke's well-known argument against the identity theory (see Kripke 1980, 150-53).

naturalisable; and so it must remain forever a mystery to us. We will never be able to provide a theory of consciousness, and consequently we will never come to understand *what* it is in terms of the natural properties from which it seems to arise and on which it seems to depend. By this measure, however, there appears to be no way of *understanding* consciousness as a natural property; at most we think of it as natural insofar as it is ontologically dependent on physical properties.

While many take the problem of consciousness seriously, that is, they suppose that unless we can understand consciousness as a physical property there is no hope of our being able to explain it scientifically, others do not think the problem is a barrier in this sense. Rather, as noted earlier, these others take the view that scientific investigations of consciousness, e.g., through neurophysiology and cognitive science, will eventually lead us to a fuller understanding of it.[4] Valerie Hardcastle succinctly describes the reasoning behind this view.[5] She contends that there is no good reason why we cannot suppose that consciousness is identical with neurophysiological properties. According to Hardcastle scepticism about the possibility of a reductive explanation of consciousness, namely, by those who take the problem of consciousness seriously, is motivated by the assumption that such explanations must allow us to understand *why* such an identity is true. But she observes that it is not the role of science to provide such explanations. She likens such a sceptic to what she calls a 'life-mysterian', that is, a person imagined to doubt that some set of biological features, e.g., reproducibility and metabolism, is identical with, i.e.,

---

[4] This general position includes a variety of views, such as those advanced by the Churchlands, Daniel Dennett, Owen Flanagan, Valerie Hardcastle, and others.

[5] See Hardcastle 1996, 7-13.

constitutive of, the property of being alive. The life-mysterian demands an epistemically satisfactory account of why this identity holds. That is, she asks that scientists show how it is impossible to conceive of being alive independent of these biological features. However, Hardcastle contends that the scientist does not have such an aim. She argues that the sceptics are analogously consciousness-mysterians – in like manner they demand that science provide an epistemically satisfactory account of why consciousness is identical with the neurological features with which it covaries. Effectively, Hardcastle reasons, these sceptics antecedently reject the possibility of naturalising consciousness, i.e., explaining it scientifically, and this reveals a fundamental difference in attitude. Moreover, she sees little chance of getting such sceptics to change their attitude. Hardcastle concludes that "[t]here are few useful conversations" (1996, 7), that is, there is little one could say to persuade the sceptics that the project of explaining consciousness in naturalistic terms is presently feasible.

Hardcastle helpfully identifies the root cause of this disagreement with respect to the problem of consciousness. However, I think that her conclusions are amiss in two very important ways. First, she is not correct to claim that being a consciousness-mysterian, i.e., a sceptic, is analogous to being a life-mysterian. The concept of consciousness, unlike concepts such as life, is distinguished by our peculiar phenomenological understanding of it, i.e., how we understand it in terms of qualia. In this sense the sceptics do not antecedently reject the possibility of naturalising consciousness as Hardcastle claims, rather they simply cannot see how consciousness understood in terms of qualia can be explained in physical terms in the way Hardcastle *assumes* they can. Second, I reject Hardcastle's assumption that it is impossible to overcome the difference in attitudes

that marks the disagreement between these disputants. There are ways to show that the sceptics' concerns are misplaced. In essence this dissertation centres on the issues surrounding these two points.

Overall, I argue that we should not endorse any metaphysical claims based on *a priori* reasoning, and which appeal to our intuitions, that contradict our scientific theories. Such a policy is liable to lead to extravagant metaphysical claims, i.e., claims which are essentially unfalsifiable. Our best policy is to accept that the ultimate authority vis-à-vis our theories quite generally is the tribunal of our senses. This is to adopt a *naturalistic* attitude to philosophical inquiry – to think of philosophy as continuous with science, as W.V. Quine urged. I argue that from this naturalistic perspective explaining consciousness in scientific terms, i.e., naturalising consciousness, does not concern our intuitive understanding of it. Therefore, insofar as consciousness is naturalisable it is so solely in terms of our naturalistic understanding of it, i.e., in terms of physiology and behaviour. Moreover, it is this naturalistic understanding of consciousness that allows us to share the concept, that is, for *us* to have a concept of consciousness at all. The concept of consciousness is nonetheless dependent on our phenomenological understanding of it. Thinking of consciousness requires thinking of it in both these ways, that is, we cannot think of it either solely in naturalistic terms or solely in phenomenological terms. These two ways of understanding consciousness are mutually dependent. They constitute a *single* concept in the sense that the one way of understanding consciousness presupposes the other. We cannot therefore think of consciousness only in terms of qualia. Indeed I argue that qualia are unindividuable, that is, there is no criterion by which we can determine whether two qualia are identical with or distinct from one another. We can still

make sense of the idea that qualia are properties, I argue, if we think of them as epistemically originating, that is, as the properties we each realise by which we apprehend things in the world. In this way we cannot expect to be able to apprehend qualia themselves, hence their unindividuability. Thinking of consciousness in this way allows us to overcome the problem of consciousness. While consciousness understood in terms of qualia is unnaturalisable, it is naturalisable in terms of our naturalistic understanding of it; and since these two understandings concern the same concept, we can say that neuroscience does offer the promise of a theory of consciousness after all.

In chapter 1 I begin by asking what is involved in naturalising a phenomenon quite generally. The answer one gives, I argue, depends on which of two different philosophical approaches one adopts. This difference in approach centres on how metaphysical considerations inform scientific claims. A scientific claim is held true on the strength of the empirical evidence for it – so long as our observations do not contradict the claim vis-à-vis its predictions within a theory, the claim is taken to be true. However, it is often assumed, in philosophical circles especially, that insofar as some claim is true it is so independently of the empirical evidence for it. By this measure no claim held true on scientific grounds in this sense guarantees that it is true as such. In other words, empirical evidence does not count as the ultimate arbiter of truth. If some scientific claim is contradicted by metaphysical considerations in particular, then we have grounds for doubting it. Let us call this the non-naturalistic attitude. Opposed to this non-naturalistic attitude is an attitude based on the assumption that there is no higher authority than the tribunal of the senses. Our best hope for holding some hypothesis about the world true is that it agrees with our observations as determined by our best scientific

theories. Thus, to the extent that a scientific claim is held true metaphysical considerations cannot undermine this. In other words, metaphysical claims cannot overrule the empirical evidence that science is founded on.

This naturalistic attitude, I contend, has an overwhelming advantage over the non-naturalistic attitude of the sceptics. The non-naturalistic attitude has intolerable consequences. It leads to metaphysical claims that cannot in principle be either confirmed or disconfirmed. I illustrate this point using two renowned historical episodes involving the denial of scientific hypotheses on metaphysical grounds, namely, René Desacartes' anti-atomism and George Berkeley's rejection of calculus. I characterise the naturalistic attitude as being guided by a precept, namely, that appeal to intuitions cannot play a determinative role in formulating our understanding of natural phenomena in terms of the theories postulated to explain them.

Then I turn my attention to the problem of naturalising consciousness and the limitations that the sceptics David Chalmers and Colin McGinn put on this possibility. Both these philosophers, I argue, adopt a non-naturalistic attitude and consequently their claims are at bottom unfounded. Chalmers hypothesises that consciousness as we realise it might arise from irreducible, i.e., fundamental, *protophenomenal* properties; where these properties are not observable and are perhaps ubiquitous. But as Chalmers admits himself, the hypothesis is empirically untestable. Therefore, it can only ever be held true on faith – there are no independent reasons for holding it true. Colin McGinn claims that we are inherently unable to explain consciousness in scientific terms. This, I suggest, is a further intolerable consequence of adopting a non-naturalistic attitude.

In chapter 2 I focus on critically evaluating a collection of now classic anti-physicalist arguments, which if successful would preclude the possibility of naturalising consciouness. The arguments I look at are Thomas Nagel's, as presented in his "What Is It Like to Be a Bat?", Frank Jackson's knowledge argument and Chalmers' updated version of the conceivability argument based on the supposed possibility of phenomenal zombies. While Nagel's argument is not strictly anti-physicalist it can be construed as such. It appears to undermine physicalism insofar as it is *impossible* to understand what it is like to be an alien creature such as a bat. I argue that there is no sense in which we can think of what it is like to be a bat. This is not something we can conceive of in any meaningful way, and so we must suspect it constitutes a pseudo-thought. Because the impossibility of understanding what it is like to be another creature is only apparent it does not threaten physicalism. Objections to both the early Jackson's knowledge argument and Chalmers' conceivability argument have been numerous. With respect to the former, I agree with Churchland's charge that Jackson equivocates between propositional knowledge concerning qualia and acquaintance with them. In the case of the latter, I agree with Peter Carruthers' objection to Chalmers' insistence on the metaphysical possibility of phenomenal zombies, namely, that the notion of property that Chalmers operates with is not naturalistic and as such should be rejected. Chalmers essentially equates properties with concepts which is contrary to the naturalistic notion of properties as existing independently of how we think of them. And this concurs with my earlier observation that Chalmers adopts a non-naturalistic attitude. All three arguments, therefore, fail to demonstrate the falsehood of physicalism. Hence, their authors do not provide any *a priori* reasons for denying the possibility of naturalising consciousness.

Chapter 3 starts with an analysis of the concept of qualia. I argue that qualia are properties that are self-evident in the sense that we do not posit their existence in virtue of needing to explain something in or about the world. This is a crucial fact about qualia. But also I argue that they are unindividuable, that is, there cannot be a criterion by which we can judge whether two qualia are identical with or distinct from one another. The perspectival nature of qualia and their self-evident existence point to their being constitutive of the subject. That is to say, we can explain the peculiar nature of qualia in these respects by understanding them not as properties the subject apprehends, but rather as properties which collectively make up the phenomenal subject, i..e, the subject understood as that which undergoes experiences. Understanding qualia in this way implies that they are not apprehensible even by the subject said to have them. Hence qualia cannot be individuated.

The unindividuability of qualia, I argue, is also suggested by Dennett. However, Dennett goes on to argue that their unindividuability implies the very concept is incoherent, and therefore we should abandon it altogther, i.e., we should repudiate qualia. I reject this conclusion, given that qualia self-evidently exist and that we can explain their unindividuability if we understand them as constitutive of the subject.

The second half of the chapter is dedicated to the objection that qualia are individuable so long as we take them to be identical with physical properties. Here I look in particular at the views of Owen Flanagan, David Papineau and Clyde Hardin. In agreement with some of Dennett's observations I contend that their argument for the individuability of qualia is unconvincing because it is ultimately question begging. They each present evidence for thinking that experiences are identical with brain states.

However, they cannot assume experiences are identical with brain states unless it is *already* assumed qualia are individuable to begin with, hence their argument is circular.

In chapter 4 I address the worry that it is precisely consciousness understood phenomenologically that demands being naturalised. If *this* is not naturalisable then it is unclear how consciousness is naturalisable. My response is to argue that this phenomenological understanding of consciousness is dependent on our naturalistic understanding of it. This naturalistic understanding concerns how we understand others to be conscious, namely, in terms of their physiology and behaviour. I argue that it seems *prima facie* possible to think of someone being conscious, as measured by behaviour and physiology, independently of being conscious phenomenologically understood in terms of realising qualia, but this is not in fact possible; that is to say, they are two aspects of a *single* concept in the sense that the one way of understanding consciousness presupposes the other. This mutual dependence follows from the fact that we cannot acquire the idea of being conscious in the one sense without the idea of being conscious in the other. So, even though consciousness is only naturalisable in terms of a naturalistic understanding of consciousness, this fact does not make our phenomenological understanding of consciousness irrelevant. Our having a naturalistic understanding of consciousness requires this other kind of understanding.

In Chapter 5 I begin by considering two worries about the position I defend, namely, that it either implies property dualism or it amounts to a pernicious form of mysterianism. In reply, I argue that my position is not threatened by dualism, whether property or substance, because I take qualia to be unindividuable. Their unindividuability means that we can neither think of *them* as physical, and therefore falling under the laws of physics,

nor as non-physical. Also, I argue that from a naturalisitic attitude there is no room for the mysterianism thought to threaten my position. The mysterianism is motivated by the assumption that there is a transcendent viewpoint from which to evaluate our theories. No such viewpoint exists for the naturalist. As epistemically originating properties we can understand how qualia are prior to the individual points of view by which we ultimately determine our theories to be true or false, i.e., through the senses. Therefore, we can understand why qualia are unnaturalisable. Hence no mystery threatens to make my position implausible.

Next I ask how it is that qualia are presumed to be real but are nevertheless unnaturalisable. Elaborating on an earlier suggestion that qualia are constitutive of the phenomenal subject I suggest that qualia are best thought of as *epistemically originating* properties, that is, not as properties we come to know, i.e., from the first-person viewpoint, but as properties *by which* we come to know things. More precisely, an epistemically originating property is a property realised by human beings, for example, in virtue of which they apprehend some aspect of the world. If a person, S, apprehends something as yellow, i.e., something seems yellow to S, say, then S realises a yellow quale. This yellow quale is epistemically originating because only in virtue of realising it does S apprehend something in the world as yellow. We know these properties to be real because without them we would not apprehend anything. What it is like for me to see red, for example, is therefore not *something* that I apprehend, rather it is that by which I apprehend. Thinking of qualia as epistemically originating properties enables us to understand how they are real but unnaturalisable.

All said, I conclude that from a naturalistic perspective the problem of consciousness can be seen to dissolve. While acknowledging that qualia are an essential aspect of our conception of consciousness, the perspective allows us to understand how they are unindividuable and consequently unnaturalisable, but still maintain that consciousness is naturalisable.

Chapter 1

Naturalism and Consciousness

That was an excellent observation. Measure the stone by the mason's rule, not the rule by the stone. But the Stoics, not applying dogmas to facts but facts to their own preconceived opinions, and forcing things to agree that do not by nature, have filled philosophy with many difficulties,...

Plutarch, 'How one may become aware of one's progress in virtue'[6]

The overall question being asked is whether it is possible to naturalise consciousness, that is, to subsume it under our scientific theories. But in order to answer this question we need a firm understanding of what it is to naturalise a phenomenon in general. That is the central undertaking of this chapter.

What will become clear in the subsequent discussion is that there are two distinct philosophical *attitudes* towards naturalising phenomena, only one of which, I argue, is naturalistic. Here it is crucial to distinguish between naturalism as an attitude or approach to philosophical inquiry and naturalism as a philosophical doctrine. The latter might be defined as the view that all phenomena are in principle naturalisable, in other words, that there are no supernatural phenomena. Naturalism as an attitude, on the other hand, is based on assuming that there is no higher authority vis-a-vis our judgments than our

[6] The Project Gutenberg's EBook Plutarch's Morals, p.78 (http://www.gutenberg.org/catalog/world/readfile?fk_files=631311& pageno=78).

senses; consequently, any conception of a phenomenon that does not fit with our empirically testable scientific theories should be revised. This is essentially epistemological naturalism. It is important to note that our concern here is strictly with natural phenomena, that is, phenomena that unequivocally concern the natural sciences. Crucially, consciousness is assumed to be a legitimate natural phenomenon in this sense. This naturalistic attitude can be usefully thought of as an approach to philosophical inquiry based on treating philosophy as continuous with science. Its most notable advocate is probably W.V.Quine. I argue that it is this attitude that provides the best approach to overcoming the problem of consciousness, i.e., the hard problem. How this attitude applies to the problem is the topic of the rest of the dissertation. Opposed to this is a non-naturalistic attitude, that is, an attitude that treats our intuitive understandings of phenomena as the grounds for evaluating the success or failure of attempts to naturalise these phenomena. It is to insist, in the case of consciousness, that any scientific theories concerning it must accommodate our intuitive understanding of it, as determined by the first-person perspective.

In this chapter, then, I focus on delineating the distinction between these philosophical attitudes. In section 1.1 I show why adopting a naturalistic approach to the problem of naturalising any phenomenon is the best policy. This is done by looking in detail at two examples in history of challenges by philosophers to attempts to naturalise phenomena on the grounds that these attempts contradict our intuitive understandings of the phenomena in question. In other words, we shall look at the non-naturalistic philosophical approach taken by these philosophers and see how and why they have ultimately failed to undermine the scientific theories they aimed to refute. In the sections

that follow this I consider the applications of this non-naturalistic approach to the problem of consciousness specifically. Here I look at the views of Chalmers and McGinn in turn. Both these philosophers embrace the *doctrine* of naturalism, that is, they hold that consciousness is in principle naturalisable. Accordingly, in order to resolve the problem of consciousness they focus on either reforming science or, in the case of McGinn, postulating a limit on our innate cognitive capacities to naturalise consciousness. I argue that the non-naturalistic attitude of both these philosophers leads to their holding extremely implausible views that contradict the naturalism they claim to endorse. My overall aim in this chapter is to show that adopting a naturalistic attitude is a promising way to overcome the problem of consciousness; and it is this attitude I shall adopt in the chapters to follow where I look in detail at the problem.

1.1 Naturalism

The philosophical naturalism I shall countenance is not a doctrine, but rather it is an attitude towards or an approach to philosophy. A naturalist in this sense is someone who treats philosophy as continuous with science. The conclusions of philosophical inquiry, by this measure, can never overrule our scientific theories. This general naturalistic attitude is exhibited by Plutarch in the quote at the beginning of the chapter, where he criticises the stoics for attempting to fit the facts to their preconceived notions of how things should be, as opposed to altering their views to fit with nature as we find it. The result, Plutarch declares, is that philosophy is filled with unnecessary difficulties. As we shall see many of the difficulties concerning how to explain consciousness are similarly

the result of not heeding Plutarch's general advice, namely, of failing to adopt a naturalistic attitude.

In general, we are never justified in revising any of our scientific theories because some metaphysical claim contradicts it. This is because insofar as philosophy aims to fix on reality, to quote Quine, "the *most* we can reasonably seek in support of an inventory and description of reality is testability of its observable consequences in the time-honored hypothetico-deductive way" (2004, 276), and testability so defined is the cornerstone of science. In other words, given that we hold our scientific theories to be true in virtue of their having been tested, and that such testing is our ultimate measure of truth, no claims derived by any other authority, in particular philosophical claims, can *falsify* these theories.[7] This continuity between science and philosophy also implies that scientific theories can inform philosophical theories.[8] We have seen a good example of this in mathematics, viewed as a science, regarding Euclidean geometry. Immanuel Kant famously assumed that Euclidean geometry is the one true description of space. He

---

[7] For example, concerning the conservation laws governing energy, charge, etc., Kenneth W. Ford remarks that "we have theoretical reason to believe that these laws are absolute...But experiment is the final arbiter. No amount of beautiful theory trumps experiment. Calling these conservation laws absolute must be as tentative as every other firm pronouncement about nature" (Ford 2004, 160).

[8] It is important to distinguish naturalism from scientism. Scientism is characterised by the attitude that assimilates philosophy with science. It can be expressed by the application of the exacting rigour of science to philosophical inquiry. More generally, scientism is the view that the methods of science provide the ultimate measure of truth. Accordingly, no philosophical claim can be determined to be true unless it is consistent with our scientific theories. Thus, science is seen as epistemically superior to philosophy. Logical postivism is scientistic for this reason. The naturalism I am urging does not assume this. Naturalism only denies that philosophical claims can be epistemically *superior* to scientific claims. So, for example, if some ethical, i.e., philosophical, claim cannot be shown to be consistent with our scientific theories the naturalist does not suppose that the claim must be false. It only denies that philosophical claims could ever overrule our scientific theories. For an excellent discussion of scientism see Bernard Williams 2006, 180-199.

thought of space and time as intuitions, i.e., space and time represent how the world *must* be presented to us; and these intuitions, he supposed, are governed by the Euclidean geometry; in other words, we cannot conceive of Euclid's fifth postulate, concerning parallel lines, as being false. However, in the nineteenth century Nikolai Lobachevski and Georg Riemann showed that we can construct a consistent geometry (indeed a set of non-Euclidean geometries) in which the fifth postulate is denied. Hence, Kant's original claim came to be seen as false. Very generally, therefore, naturalism can be described as the denial that there can be a more secure independent philosophical viewpoint from which science and other knowledge claims can be evaluated.

Because naturalism is defined by practice, i.e., as an approach to doing philosophy, the best way to understand its justification is to consider historical examples of philosophers' attempts to circumscribe scientific inquiry. How these attempts have failed illustrates why naturalism is *good* practice. We shall look at two cases. The first example concerns Descartes' scepticism about experiments concerning vacua; where Descartes argued on metaphysical grounds that it is impossible for there to be empty space. The second concerns Berkeley's sceptical attack on infinitesimal calculus, despite calculus's obvious fecundity as a tool of the physical sciences. These examples will be presented in what at first might seem like excessive detail. But this detail is necessary because of the central role of these examples in the overall argument of this chapter.

René Descartes famously tried to apply a mathematical method to science, as outlined in his *Rules for the Direction of the Mind*. Very briefly, assuming an epistemological foundationalism he thought that scientific knowledge (*scientia*) can be acquired by a bottom-up process. This process involves first determining those truths we can hold with

certainty – these being intuitively self-evident according to Descartes – and these revealed truths are then used to deduce how things are in the world in a very general sense, e.g., in terms of laws of nature, or as he calls it "the truth of things" (1985, 15). Thus, Descartes conceived of science as being an axiomatic system analogous to Euclidean geometry. The wrongheadedness of this approach is illustrated by his disagreement with Blaise Pascal and other pioneering experimental physicists regarding their attempts to produce a vacuum. Two of Descartes' metaphysical claims are relevant to this debate.

First, Descartes held that space is necessarily instantiated by corporeal substance (as distinguished from non-spatial mental substance). In *Principles of Philosophy* he stated: "It is easy for us to recognize that the extension constituting the nature of a body is exactly the same as that constituting the nature of space" (ibid, 227). He seems to have taken this to be self-evidently true, appealing to our *intuitive* understanding of space. He acknowledged that we can conceive of space independently of any *particular* body (*ibid*, 228). For example, we understand a stone as occupying some space, but we can imagine the same space remaining after the stone is removed from that place. However, Descartes insisted that this space, formerly occupied by the stone, still exists in its absence insofar as it is instantiated by some matter, e.g., air. A vacuum, on the other hand, is understood as space independent of any corporeal substance at all. Descartes thought that this is an incoherent idea. Thus, if in a sealed vessel, say, it were possible to remove all matter such that it could be said to be a void, then, according to Descartes, we would have to conclude that there is no distance between the sides of the vessel.

The second metaphysical claim relevant to the debate is his denial of atomism. The impossibility of atoms follows from space being necessarily instantiated by corporeal substance. Any body is continuously extended, i.e., there are no gaps, since gaps or voids occupy no space. Hence, every body must be perfectly dense in this sense. Perfect density of bodies requires, according to Descartes, that they are infinitely or indefinitely divisible – essentially, in order to ensure that no gaps exist (*ibid*, 239). Atoms, thought of as *finite* packets of discrete matter, therefore cannot in principle exist. Descartes thought that these claims about nature are logically entailed by self-evident truths and so they are indubitable – they amount to genuine knowledge.

Many of Descartes' contemporaries, however, did endorse some sort of atomism. Atomism suggests the possibility of vacua, at least to the extent that it is most plausible to assume that there is empty space between atoms. Accordingly, attempts were made to create a vacuum. One such early experiment was by Evangelista Torricelli, a student of Galileo, who used a long glass tube with mercury and a bowl containing water with a layer of mercury at the bottom. He carefully covered the one open end of the tube, turned it downward and immersed it in the mercury at the bottom of the bowl. After uncovering the end he observed the mercury drop part way down the tube leaving behind nothing, since no gases were observed to bubble up to fill the empty space.[9] Intrigued by reports about this experiment Blaise Pascal conducted a very similar test on top of the mountain Puy-de-Dôme some years later. He reasoned that the lower atmospheric pressure at the greater altitude would result in the mercury dropping farther down the tube than at close

[9] A detailed account of these experiments and the ensuing debate is given by Daniel Garber in his book *Descartes' Metaphysical Physics* (see especially Garber 1992, 136-143).

to sea level, which indeed it did. There was nothing to show at least that a vacuum was not really present in the top of the tube.

Descartes' reaction to these experiments was interesting. Despite his holding that vacua are in principle impossible, he took the relatively agnostic position that he might accept the presence of a vacuum in the tube if enough evidence was presented to him. Pascal never gave him a report detailing such evidence, so Descartes remained unconvinced. Moreover, Descartes argued that the space at the top of the tube was in fact filled with an infinitely fine ether that occupies any potential gaps, including pores in the glass through which it permeated the inside of the tube. Thus, he believed that there was no vacuum present.

His anti-atomism is a philosophical theory in virtue of the fact that it is justified *a priori*, that is, solely in terms of an intuitive understanding of space and material bodies. At no point did he concern himself with how well his theory fitted with experience. His thought would seem to have been that what we observe *must* agree with his theory because it is deduced from certain knowledge, i.e., that which cannot in principle be doubted, hence his overall lack of concern with the fact that the experiment's suggestion that vacua are possible. More generally, because Descartes did not propose his anti-atomism in order to explain any phenomena he did not seriously consider its predictive efficacy. However, he seemed to be cognisant of the fact that his theory should be able to explain phenomena, hence his attempt to explain the *appearance* of a vacuum in terms of the space at the top of the tube really being instantiated by an infinitely fine ether.

Descartes' argument against the possibility of vacua, in light of this contrary experimental evidence, did not convince these early scientists to give up atomism. The

reason is that, unlike Descartes, their main interest did not appear to be to develop a theory that accommodated our intuitions about space and matter, however strong these intuitions might be. Instead, their concern was more precisely to see how well atomism agreed with the empirical evidence, hence their interest in the experiment. Thus, their aim was crucially different from Descartes'.[10] The sort of *a priori* justifications Descartes had for his claims were irrelevant to these scientists. This point can be generalised: The purpose of a scientific theory is to explain phenomena, and as such a theory is held true so long as it succeeds in this, regardless of whether or not it conflicts with our intuitions. That is to say, if a scientific theory is very successful at explaining relevant phenomena and is predictively powerful but nonetheless conflicts with our intuitions, then so much the worse for our intuitions. Insofar as philosophical claims are ultimately justified only in terms of our intuitions, therefore, they cannot undermine our scientific theories.

That said, sometimes our intuitions seem so strong that when a theory contradicts them we cannot help but think that the theory is in some sense *ad hoc*, that is, we feel that the price of accepting the theory is that we must ignore its inability to accommodate some of our most basic beliefs. In this way a theory can seem mysterious to us despite its strengths and usefulness. In such circumstances the scientists' response is to urge us to let go of the intuitions in question. A good example of this sort of assault on our intuitions and the above type of response concerns the advent of infinitesimal calculus, as originally developed in its modern formulation by Gottfried Leibniz and Isaac Newton independently of each other in the latter half of the seventeenth century.

---

[10] That is not to suggest that Descartes' thinking was primitive in comparison to some of his contemporaries. No one at that time could be described as having a modern scientific attitude, but in *this* instance Descartes' thoroughly metaphysical approach was problematic.

Clearly calculus, as a branch of mathematics, is not directly concerned with observation in the same way as atomism. The theories of calculus cannot be confirmed or disconfirmed by experimentation. However, the theories of calculus are nonetheless indispensable to natural sciences, especially physics. Indeed, without calculus the present day natural sciences would not exist. Thus, the consequences of denying the truth of calculus quite generally would be pretty dire – it would effectively rob us of the principal justification for a great many of the claims made by the natural sciences. The usefulness and applicability of modern calculus to the sciences was obvious to its creators and many of their contemporaries.

The calculus of Leibniz and of Newton originated from a tradition dating back to the Ancients of the analysis of space and motion in terms of its parts. Pythagoras, or his followers, employed what was called a process of exhaustion to calculate the areas of various shapes for example. That is to provide closer and closer approximations to the areas of different shapes by 'filling in' some area with a greater number of ever smaller shapes with calculable areas, e.g., squares and triangles. It was the development of Cartesian coordinates, many years later, that allowed both Newton and Leibniz to analyse space and time in this way more effectively. For example, the motion of a body can be described by a curve plotting time elapsed, t, against distance traveled, s. Moreover, the tangent at any point along such a curve gives us the velocity of the body at that instant. Again we can think of this body as traveling a constant velocity for a brief period, Δt, covering a distance Δs and then instantaneously changing velocity. This approximation to the continuous motion of a body could describe a discrete curve, as it were, comprising segments of straight line joined end to end. The velocity of the body over each brief

period is equal to the tangent of the particular segment of straight line, i.e., $\Delta s/\Delta t$. The *continuous* motion of a body is therefore represented by the aggregate of such segments over infinitesimally brief periods of time, i.e., at each instant; where an infinitesimal quantity in general is conceived as some indivisibly small quantity that is nonetheless greater than zero. At each instant in the motion of a body, therefore, its velocity is $\Delta s/\Delta t$ where Δs and Δt are thought of as infinitesimal. In the now standard Lebnizian notation this ratio, known as a derivative, is written as ds/dt and it represents the velocity of a body in continuous motion at a particular instant, i.e., at the limit as Δs and Δt *approach* zero.

Both Leibniz and Newton contended that this limit should not be thought of as ever being reached, since at that point the derivative would absurdly be the indeterminate ratio 0/0. However, insofar as we think of motion as continuous the differentials ds and dt at an instant, represented as a point on a continuous curve, cannot be distinguished from those of their neighbouring instant except by supposing them to be *at* their limit, rather than merely approaching it. This is essentially the problem concerning Zeno's paradox as described in terms of the motion of an arrow. Zeno noted that an arrow's motion in flight is infinitely divisible, assuming that motion is continuous, such that it ultimately comprises a sequence of *distinct* infinitesimal instants, each effectively being of no duration. But, if at each instant no time elapses then the arrow does not move. Hence, the arrow's motion thought of as the aggregate of these infinitesimal instants can only add up to zero; that is to say, the arrow is in fact motionless, a similarly absurd result. Newton's characterisation of the differentials, or 'fluxions' as he called them, does not help to overcome this difficulty. He described fluxions as evanescent quantities, that is, thought

of in terms of the senses, quantities that begin to vanish at the very moment that they arise; consequently, he explained, "those ultimate ratios with which such quantities vanish are not truly the ratios of the ultimate quantities but the limits to which the ratio of quantities, decreasing without end, always converge" (Boyer 1959, 216). But these *infinitesimal differentials*, of which derivatives are ratios, remain mysterious.

It is this difficulty that George Berkeley pointed to in criticising the methods of calculus in his paper 'The Analyst' (1734). He wrote: "Now to conceive a quantity infinitely small, that is, infinitely less than any sensible or imaginable quantity, or the least finite magnitude, is, I confess, above my capacity" (Jesseph 1992, 167). But his most powerful criticism of the methods of calculus concerns Newton's and Leibniz's demonstration that for any function of the kind $y = x^n$ its derivative, i.e., dy/dx, is nx^{n-1}. The demonstration is briefly along the following lines: The derivative represents the ratio of some increment of y to an increment of x, let the increment of x be o. The increment of y, therefore, is $(x + o)^n - x^n$. This multiplies out to be $x^n + nox^{n-1} + (n^2 - n)o^2x^{n-2}/2 + \ldots - x^n$. The ratio is arrived at by dividing this expression by the increment o, which gives us $nx^{n-1} + (n^2 - n)ox^{n-2}/2 + \ldots$. At this point, however, it is reasoned that as o approaches zero this expression approaches nx^{n-1} at its limit. Berkeley complained that this result is only arrived at if o is first assumed to be a quantity, i.e., as a divisor, and then afterwards assumed to have no magnitude at all, i.e., to be zero. These assumptions, Berkeley noted, are contradictory. And he warned that "no just conclusion can be directly drawn from two inconsistent suppositions" (*ibid*, 175). This criticism applies to all types of functions given that the same method is used to derive the expression of the derivative for any function.

What Berkeley's criticisms highlight is the difficulty in conceiving of the infinitesimal. As Carl Boyer notes, Leibniz conceded that "one could not prove or disprove the existence of infinitely small quantities," but nonetheless he believed that his calculus was justified on the basis of its usefulness (Boyer 1959, 217). Moreover, at one point Leibniz described infinitesimals as useful fictions, but he insisted that the truths concerning them hold in virtue of the fact that the notion of continuity requires them in some sense, that is, insofar as we think of space and motion, say, as continuous we must think of them as really being infinitely divisible (see *ibid*, 219). The problem of infinitesimals, then, stymied the defenders of calculus. While calculus's overwhelming usefulness made it an indispensable analytical tool of the physical sciences, the lack of any rigorous demonstrations of its theorems meant that why these theorems are true was a mystery.

Eventually the root of the problem was diagnosed as being our intuitive understanding of continuity in spatiotemporal terms. To the extent that we think of continuity in such terms we cannot fully make sense of the notion of the infinitesimal. If, like Zeno, we think of an arrow's motion as continuous in this way, then we understand that no matter how many times we divide up the period of its flight it is always possible to divide it more. If we were to suppose, on the other hand, that such divisions really reach a limit, it is hard to understand how each infinitesimal period, i.e., each instant, can have any magnitude, since if it did we could imagine its being divisible still. The solution to the problem, therefore, is to abandon thinking of continuity in spatiotemporal terms. Mathematicians began to construe continuity in analytical terms, that is, strictly as a relation between numbers. One might say that they define continuity operationally. Thus,

the infinitesimal is informally thought of as an arbitrarily small quantity, that is, as small as one likes. In this way a function, $f(x)$, represented by a curve is said to be continuous at $x = a$ if for any value of $f(x)$ approaching this value as a limit, i.e., $f(a)$, there is some positive number ε, i.e., $\varepsilon > 0$, such that $| f(x) - f(a) | < \varepsilon$, then there is always some number δ, where similarly $\delta > 0$, such that $| x - a | < \delta$. [11] This definition of continuity in terms of limit does not invoke spatiotemporal concepts, it is purely analytical. Calculus conceived in this thoroughly formalistic manner avoids the difficulties pointed out by Berkeley; expunged of any empirical content in this way the infinitesimal, specifically, is no longer thought of as a magnitude.

With the example of Descartes' anti-atomism we saw that appeal to an intuitive understanding of space to justify such a theory is denied; this approach is backwards in the sense that science aims to develop theories that agree with phenomena rather than aiming to explain phenomena in terms of theories we find intuitively compelling. Similarly, with the example of infintesimal calculus, we saw that mathematicians aim to develop theories that are ultimately self-consistent, regardless of whether these theories fit with our intuitive understanding of the concepts involved, e.g., continuity understood as that which is unbroken, like a chalk line drawn on a board. Accordingly, undergirding the sciences in general is a fundamental attitude that might be expressed by the following precept: How the world is, *as measured by our experiences*, should govern how we think of it. Thus, how we think of the world should not ultimately be determined by our intuitive conceptions of it, however compelling these conceptions might be. That is not to say that our intuitions play no role in constructing our theories, only that at bottom

[11] This definition originates with Karl Weierstrass (1815-1897) in virtue of the earlier insights provided by Augustin Louis Cauchy (1789-1857) in particular.

whatever theory is constructed it must agree with experience, i.e., be empirically confirmable either directly of indirectly. In terms of naturalism, an entity (object, property, etc.) is understood to be *natural* according to how it is conceived by our best scientific theories. So, for example, life as a natural property is understood strictly in terms of what our biological theories – broadly defined – take it to be, e.g., in terms of reproduction, metabolic processes, etc. Thus, to think of life as an irreducible force, such as élan vital, instantiated by certain complex objects is not to understand life as natural.

These historical examples illustrate how metaphysical considerations based on reasoning that appeals to our intuitions cannot overrule scientific claims. The practice of science in this respect is not thwarted by philosophical concerns. Scientists, in other words, do not have to change their practices in order to accommodate philosophical doubts about them. The philosopher who takes this on board essentially adopts a naturalistic attitude. Let us next consider how exactly this naturalistic perspective informs our understanding of consciousness.

1.2 Consciousness as a Natural Property

I stated that naturalism is not a doctrine but rather it is an attitude. And this fact has the important consequence that it cannot be said to be true, or indeed false. At most we can say that naturalism is good, or conversely, that not to take a naturalistic approach to philosophical inquiry is unpropitious. Denying naturalism in this sense is to assume that our scientific theories must accommodate certain metaphysical assumptions that are justified on an *a priori* basis, i.e., by appeal to intuitions alone. Adopting this attitude leads the philosopher to suppose that there is some aspect of the world that is effectively

closed to scientific inquiry, at least as it is presently conceived or practised. We saw by way of the two historical examples that the problem with such a stance is that it is antithetical to the basic project of science, namely, to build an understanding of the world that best fits with our experience of it. The scientist is not interested in how we think the world *must be*, rather she is concerned with how we should think of the world according to the evidence available to us by our senses. Here it is not being suggested that the scientific theories we propose are *determined* by our observations, such that scientists should reject concluding, for example, that 'moderate-sized specimens of dry goods'[12] such as tables and chairs are almost wholly composed of empty space, since this is contrary to how such items appear to us. Rather, the point is that whatever theories and claims we make about the world they must ultimately be accountable to our senses. That medium-sized objects are composed of atoms and large amounts of empty space around them is ultimately consistent with our observations in this sense. In the case of Descartes' anti-atomism, he seemed to presume that we cannot doubt that vacua are impossible in principle irrespective of the evidence to the contrary. Likewise, Berkeley supposed that the very notion of continuity precludes us from being able to justify calculus, despite its obvious fecundity as an analytical tool of the physical sciences. In this latter case we saw that Berkeley did not appreciate the fact that how we think of continuity is not fixed. What is fixed is the demand that mathematics be self-consistent.

Where this non-naturalistic streak is at its strongest in philosophy is with respect to consciousness. This reflects the fact that it seems impossible for us not to think of consciousness phenomenologically, that is, as we are acquainted with it from the first-person point of view; and yet this conception strikes us as being beyond description in

[12] This phrase is borrowed from J.L.Austin, (see 1962, 8).

physical terms, a point, as we shall see in the next chapter, astutely made by Thomas Nagel. But does this fact imply that consciousness, as a property, cannot be understood other than phenomenologically? To assume this seems too strong, since we understand other types of creatures to be conscious and attribute consciousness to them, even though we have no access to their points of view. This suggests at least that we *also* understand consciousness as a straightforwardly observable property.[13] Consciousness in some sense seems to be amenable to investigation from the third-person point of view. It seems more than plausible to suppose that consciousness can be understood as a natural property as defined above, at least to some degree. That is to say, in their investigations of conscious experiences neuroscientists are not changing the subject. This is not to presume that conscious states are identical with brain states such that the study of consciousness is straightforwardly a matter of empirical investigation; rather, it is to suppose that consciousness is understood in two ways, i.e., from the first- and third-person perspectives, and these *ways of understanding* are mutually dependent on one another. How they are mutually dependent is discussed in detail in chapter 4.

1.2.1 Chalmers' Naturalistic Dualism

It is the difference between a naturalistic and a non-naturalistic attitude that Valerie Hardcastle recognises as the root of disagreement between physicalists and anti-physicalists, or 'sceptics' as she calls them. And because such disagreement is a matter of attitude, she concludes that "this difference is not something that further discussion or

---

[13] This claim is not wholly uncontentious. Peter Carruthers, for example, has argued that insofar as we can distinguish between conscious (what-it-is-like) experiences and non-conscious ones, it is unclear whether other creatures, or 'brutes' as he calls them, have the former type of experiences (see Carruthers 1989).

argumentation can overcome" (1996, 7). As a committed naturalist she protests that the non-naturalistic approach of the sceptics has unfortunate consequences. She takes as her example Chalmers' views. Specifically, Hardcastle considers Chalmers' claim that consciousness is a fundamental property, a brute fact about the world. Its being a property of things is analogous to gravitational attraction in the sense that we do not suppose its existence can be explained in simpler terms. Hardcastle finds this way of thinking of consciousness counterintuitive. She observes that if one claims that consciousness is simply a brute fact in this sense "then one is *prima facie* operating with a perverse metaphysics" (ibid, 9). Chalmers suggests that the phenomenological quality of experience is an aspect of information. Hardcastle points out, however, that "not all information has a phenomenal edge, insofar as we know quite a bit of our information processing [i.e., in our brains] is carried out *un*consciously" (*ibid*). And she notes that to explain this fact Chalmers seems forced to suppose either that all our information processing in the brain is conscious, but we do not realise it, or that no information processing is occurring in those cases where the brain processing is unconscious. Neither of these options is at all plausible.

Chalmers' position here seems relevantly analogous to Descartes' with respect to his defence of the claim that vacua are impossible inspite of evidence to the contrary. Like Descartes vis-a-vis his rejection of atomism Chalmers can offer a plausible defence of his denial that consciousness is a biological property. Just as Descartes defended his position from premisses that were not formed with any intention to fit with the empirical facts, so Chalmers dismisses such considerations when forming his basic assumption, namely, that consciousness is a fundamental property. Chalmers' theory is not formed with the

intention of explaining any empirical facts since he assumes, albeit quite reasonably, that phenomenological facts are not empirical. Indeed, he acknowledges that "[b]ecause consciousness is not directly observable in experimental contexts, we cannot simply run experiments measuring the experiences that are associated with various physical processes, thereby confirming or disconfirming various psychophysical hypotheses" (1996, 215). And here he admits that his theory of consciousness is empirically untestable. The problem is, however, that if no empirical evidence can ultimately count in favour of or against his theory, then it is unfalsifiable. This makes the theory logically independent of science – whether it is true or false makes no difference with respect to its consistency with our scientific theories quite generally.

Chalmers is aware of this difficulty with his view. His response is to argue that his theory is falsifiable by a process of inference to the best explanation (see Chalmers 1996, 215-218). He argues that while his theory is untestable, testing is not the only way to evaluate a theory, i.e., to decide whether it is true or false. In addition, he supposes, we can hold a theory true if it offers the most plausible explanation we can arrive at. However, if, as he admits, his theory is untestable in principle, in what sense can we hold it true? To suppose a theory can be thought of as true *solely* on the grounds that it is plausible, despite being untestable, is untenable. Imagine the following relevantly analogous scenario. Two centuries ago a man, Mr Smith, is reported missing after going for a walk near his home. He was never seen again either dead or alive. There are, let us suppose, several explanations about how and why Mr Smith disappeared, of varying degrees of plausibility. However, we cannot hold what seems like the most plausible explanation as true, given that none of the explanations are testable. That is not to deny

that one of the explanations might *be* true. In general, we hold our theories to be true insofar as they are at least testable in principle.[14] Therefore, the most plausible explanation of Mr Smith's disappearance is at best a conjecture. Here we are assuming, for argument's sake, that his disappearance is so remote in time that no evidence could have survived. Chalmers' theory cannot be held true, however plausible, for the same kind of reasons.

One might object that inasmuch as truth is evidence transcendent one could assume that one of the explanations of Mr Smith's disappearance is true, despite our being unable to determine which. Therefore, the most plausible explanation could be held as true. But, to hold an explanation or theory as true entails offering justification for *its* being so. An explanation's plausibility does not on its own justify our believing it is true – it is not sufficient for justification in this sense. No matter how convinced one might be of a particular explanation, e.g., about Mr Smith's disappearance, this alone is never enough to enable us to hold it true. The strength of one's conviction is irrelevant with respect to its truth; that is to say, how strongly you believe something to be true obviously does not make it so.[15]

---

[14] This is not to affirm the verificationist principle of truth, namely, that something *is* true if and only if it is empirically verifiable. Rather, the claim is that a theory (about natural phenomena) can be *held* true if and only if it is testable; where to hold a theory true entails *acting* as if the theory is true. So, while one could act as if one explanation of Smith's disappearance is true, nothing about the facts in the world agree with, or best fit with, acting in this way rather than in some other way, i.e., as if some other explanation about Smith were true.

[15] Contemporary physics provides us with a dramatic example illustrating this point, namely, the case of superstring theory. Since the 1980's a great deal of research in theoretical physics has focused on this theory. It promises to unify the theory of relativity with quantum theory, that is, to incorporate under this one single theory the force of gravity and the three forces found at the quantum level, i.e., electromagnetic force and the weak and strong nuclear forces. However, the size of strings is so small (estimated between 10^{-16} and 10^{-33} cm) that they are unobservable – either directly or indirectly. But

More generally, our scientific theories and explanations are not held true simply because of how neatly we feel they fit with our present beliefs, and which are therefore judged to be most plausible; rather, they are held true to the extent that they are predictively successful, i.e., how well they tell us what will happen. Certainly one might think that Einstein's general theory of relativity, for example, is held true because it offers the most plausible explanation of, say, Mercury's shifting perihelion. However, this theory is held true because of its predictive power rather than its offering the most plausible explanation. Its competing theory, i.e., Newton's, fails to *predict* this phenomenon. Einstein's theory, by contrast, predicts the curvature of space which can explain the phenomenon. In this sense what interests us in a theory is that it resists being falsified, and this is a measure of its predictive power. And here Chalmers' theory of consciousness, however plausible it might be, predicts nothing – it is predictively inert and hence unfalsifiable. Its logical independence implies that holding it true has no consequences. I am not suggesting that inference to the best explanation is illegitimate. It is a legitimate way of deciding which theory among a set is best to hold true. But we hold the chosen theory true *on condition* that it is testable, i.e., to the extent that it has predictive power.

It might be contended that if one countenances some sort of epistemic holism, e.g., Quine's, then one can suppose that so long as a theory is consistent with the rest of our

---

more importantly, string theorists have never been able to devise an experiment to test the theory (or some version thereof). In other words, as it stands string theory is unfalsifiable. Indeed, many physicists suspect that the only hope of testing it requires reproducing the extreme conditions thought to exist very soon after the big bang, and this seems impossible. All told the present status of string theory is precarious. The theory is mathemetically sound and is consistent with physics; in this sense it is a very plausible theory. The difficulty is that it predicts nothing about the observable physical world.

scientific theories we can hold it true, despite its not being directly testable. So, the idea is that while the hypotheses of some theory, call it T_1, are not testable, so long as T_1 can be shown to be consistent with the rest of our theories, T_2, T_3, ..., T_n, then we can hold T_1 as true, where these other theories are testable. Accordingly, one might suppose that while the hypotheses of T_1 are not themselves directly testable they are indirectly so, in virtue of T_1's consistency with the other empirically confirmable theories. But this defence of Chalmers' position will not work for the simple reason that the theory he postulates is logically independent of the rest of science. Even if we hold the hypotheses of Chalmers' theory as false still the other theories can be assumed as true. Therefore, no hypotheses of his theory can be justified by the rest of science in this way.

But again, Chalmers admits that his theory is not hard science in that it lacks the "empirical credentials" of other sciences (*ibid*, 218). Moreover, the social sciences are similarly vulnerable to the empirical untestability of their theories. Therefore, it might be contended that Chalmers' theory is properly scientific despite its untestability. But given that his theory concerns the relations between consciousness as a basic property and the basic physical properties, such untestability does not sit well. The stated goal of his theory is to naturalise consciousness, that is, to incorporate it into the natural sciences. The theory lacks the autonomy afforded to social scientific theories in this regard. We do not expect social phenomena to be naturalised in this way, i.e., explained in physical terms. For example, Darwin provided us with a powerful theory , i.e., natural selection, to explain how species originate. But the *concepts* involved in this theory, e.g., species, reproduction, survival, and so on, are not reducible to basic physical properties. One can

---

Therefore, strictly speaking, it does not presently qualify as a theory of physics – it remains a plausible conjecture (see Lee Smolin, pp. 177-199).

fully grasp these concepts without deferring to such concepts as charge, electron, etc. The theory of natural selection and physics are independent of each other in this respect. We do not suppose there are any laws relating the phenomena of these two theories. By contrast, Chalmers assumes there are laws relating consciousness to basic physical properties.

1.3 Varieties of Naturalism

It might seem odd to characterise Chalmers' view as being antithetical to naturalism given that he calls it 'naturalistic dualism'. Indeed, many of those sceptical about the prospect for science, as it is practised, providing a theory of consciousness call their views naturalistic nonetheless.[16] There is at work here an understanding of naturalism that is quite different from that I have outlined. There are several varieties of naturalism. Chalmers claims that his view is naturalistic because "it posits that everything is a consequence of a network of basic properties and laws, and because it is compatible with all the results of contemporary science" (1996, 128). However, as we have seen, this compatibility is trivial since his theory is logically independent of science. As well McGinn calls his view 'non-contructive naturalism'. It is naturalistic, he tells us, because it denies that consciousness is supernatural, that is to say, "it must be in virtue of *some* natural property of the brain that organisms are conscious" (McGinn 1991, 6). McGinn and Chalmers think that consciousness is naturalisable, i.e., can be incorporated into the sciences, and consequently that it is a natural property. However, for Chalmers consciousness is naturalisable only if our natural sciences are radically reappraised, or

reformed, by thinking of consciousness as a fundamental property. This is despite the fact that his theory is logically independent of the rest of science. In other words, his reforms are unjustified. These reforms assume there can be *scientific* hypotheses that are in principle empirically untestable, either directly or indirectly. But any such hypotheses cannot of course be scientific, since the natural sciences are predicated on their empirical testability. And for McGinn, while consciousness is naturalisable we cannot formulate the theories necessary to understand how it is natural, since, as we shall see, he contends that our innate cognitive limitations preclude us from ever being able to do this. Therefore, for both of them, consciousness cannot be understood as a natural property *according to science as it is (currently) practised by us.*

These philosophers, then, suppose that metaphysical considerations concerning consciousness can influence how we view science. That is, they assume that we can appraise our sciences from an independent philosophical viewpoint. This is to adopt a non-naturalistic attitude in terms of the naturalism I have outlined using the historical examples concerning vacua and calculus. Unless stated otherwise, by 'naturalism' I refer to the version I have countenanced. Again, this involves treating philosophy as continuous with science, not independent of it. To adopt a *naturalistic* attitude, therefore, requires a willingness to give up our intuitive understandings of phenomena. Science only starts when we take this stance, which can be described by the precept I mentioned earlier, namely, that how the world is, as measured by our experiences, should ultimately govern how we think of it. That is because there is no higher tribunal by which we can judge our claims about the world than our senses.

---

[16] I have in mind here Strawson's naturalised Cartesianism as expounded in his *Mental Reality* (1992), as well as Colin McGinn's so-called non-constructive naturalism

The appeal of deferring to our intuitions can be very strong in philosophy, given its centrality in philosophical reasoning quite generally. It is this reliance on intuitive understandings of phenomena that Dennett also criticises. He cites a project by AI researcher Patrick Hayes to formalise what Hayes calls our naïve physics, i.e., our everyday understanding of how the physical world operates, e.g., unsupported objects invariably fall towards the ground etc. (Dennett 2005, 31-35). The expectations that constitute our naïve physics are internalised by us thoroughly; they play an indispensable role in our everyday actions, such as quickly moving away from a glass full of water as it tips over, expecting the water to spill out. Of course, the 'theories' of naïve physics are false. For example, according to its theories liquid in a drinking straw will fall out, whereas in fact it remains inside. Dennett notes that a similar formalisation could be done for folk psychology, and likewise its theories would be false. Folk psychology fails to predict such anomalous phenomena as blindsight and prosopagnosia (inability to recognise faces). Such naïve theories are false because they are ultimately determined by our intuitive understandings, i.e., how we think the world should operate, rather than how it does. We can interpret certain philosophers of mind as attempting to formalise rigorously our folk psychology, that is, to explain how our intuitive understandings of such mental phenomena as consciousness are true. Some such philosophers at least have, as Dennett describes it, "proceeded as if the deliverances of their brute intuitions were not just *axiomatic-for-the-sake-of-the-project* but *true*, and moreover, somehow inviolable" (*ibid*, 34). Dennett notes William Lycan's quote of Nagel's appeal to the primacy of intuition in this regard in his book *Mortal Questions* where Nagel asserts that "I believe one should trust problems over solutions, intuition over arguments"(*ibid*, 22, n. 18).

---

and Chalmers' naturalistic dualism.

While scientists are often pleased that their results are counterintuitive, for some philosophers of mind a claim's being counterintuitive is taken to show that it is false (*ibid*, 34). Here it is important to add, I think, that what encourages this kind of deference to intuition is that the resultant theories are often not shown to be false, as we have seen with Chalmers.

The theory of consciousness that Chalmers advances is by his own admission untestable. He claims that we can hold it true in virtue of its plausibility. I have argued, however, that its untestability makes his theory unfalsifiable. Consequently, the theory is not constrained by its empirical viability – it has none. Not surprisingly, therefore, the hypotheses he postulates in relation to his theory are at best speculative and at worst metaphysically extravagant. He is led to propose, for example, that consciousness is not a chauvinistic biological property of certain complex organisms, but it is possibly a ubiquitous feature of countless other objects, ranging from coffee makers to cars. All this is a direct result of Chalmers' non-naturalistic approach. He assumes that how we think of consciousness in phenomenological terms must be accommodated by our sciences, and insofar as it cannot we must reform our scientific practices. Accordingly, he recommends that consciousness ought to be thought of as a fundamental property of the world alongside such physical properties as length and mass. As we shall see, this idea is very problematic (see section 2.3). But the non-naturalistic approach of McGinn does not seem to lead to the sort of extravagant metaphysics that Chalmers countenances. McGinn does not prescribe any reforms of our scientific practices, and this would suggest a more plausible version of naturalism. Let us look at McGinn's view in more detail.

1.4 Consciousness and Cognitive Closure

The central premiss of McGinn's argument is that there must be some property, P, of the brain that causes or generates consciousness given that consciousness is naturalisable in principle. To suppose instead that our being conscious is just a brute fact, that is, to suppose that consciousness arises from the brain causelessly, is to treat it as miraculous (1991, 6). However, McGinn argues that P is not an ordinary physical property, that is, a *perceptible* property, e.g., a neurological property of the brain, since, as he puts it, "[n]o matter what recondite property we could see to be instantiated in the brain we should always be baffled about how it could give rise to consciousness" (*ibid*, 11). And P is precisely the property the knowledge of which would allow us to connect conscious states to physical brain states. In other words, McGinn thinks of P as what might be described as a 'psychophysical' property, that is, a property that we can understand as relating to both consciousness and the brain. P is the 'missing link' that is needed for us to be able to subsume consciousness under our scientific theories, i.e. to naturalise consciousness.

In order to explain consciousness therefore, according to McGinn, we first have to apprehend P. He recognises two cognitive faculties by which we might be able to do this, namely, perception and introspection, but he argues that neither of these faculties allows us to apprehend P. Introspection gives us each direct access to the properties of consciousness, e.g., we can apprehend the experience of pain, say, but it does not let us see how this experience is related to the neurological processes that are correlated with it (*ibid*, 8). Thus, we cannot form a conception of P by introspection; P, after all, is a psycho*physical* property and introspection only allows us to apprehend properties of

consciousness *simpliciter*. Encouraged by neuroscience, however, we might suppose that we can grasp P through our senses, i.e., perceptually. The problem with this route is that nothing we perceive in the brain specifically has to be thought of as being related to consciousness. McGinn argues that nothing we perceive "could ever convince us that we have located the intelligible nexus we seek" (*ibid*, 11).

Ultimately McGinn argues that P and consciousness itself are essentially non-spatial properties. Accordingly, they are best thought of as theoretical given that their non-spatiality makes them imperceptible in principle, or unobservable. This fact, he thinks, does not rule them out from being natural since quantum theory posits theoretical entities in this sense, entities that are clearly regarded as natural (*ibid*, 12). Assuming that P is theoretical it seems plausible that we can at least deduce its existence by inference to the best explanation. That is, so long as the theory that best fits the empirical data, and which is thereby held as true, posits P, it seems possible to construct a suitable psychophysical theory that can explain consciousness in terms of P, despite P's being imperceptible in principle. McGinn, however, denies that this is a possibility. Our scientific theories, McGinn suggests, employ models for those properties it aims to explain that are essentially analogical extensions of perceptible macrophysical properties. He gives the example of how our concept of molecule is based on its representation in terms of such macrophysical objects (*ibid*, 13). That is, we think of the molecule as being analogous to spheres representing atoms, rods connecting them, and so on. Such explanations, therefore, require a homogeneity between such macrophysical objects and the objects or properties concerning a theory. It is this homogeneity that a psychophysical theory lacks. Properties of consciousness and the physical properties we want to relate them to,

McGinn asserts, are heterogeneous in this respect. The models used for our physical theories are useless with respect to constructing a suitable psychophysical theory that concerns explaining non-spatial properties of consciousness, i.e., properties that are disanalogous to our perceptually determined *spatial* models.

Consider an ordinary example of a physical explanation: We explain an object's being black in terms of the microphysical properties of the object that cause it to absorb light. We can understand these microphysical properties as being 'blackening', i.e., as fulfilling the functional role of making the object appear black. This explanation works, i.e., is satisfactory, to the extent that we grasp how these microphysical properties fulfil this functional role. This we can do only if our models of these microphysical properties are analogical extensions of perceptible entities. Such spatial models will be of no explanatory efficacy vis-à-vis a psychophysical theory, that is, they could not help us to understand non-spatial properties of consciousness. Thus, because P is imperceptible, McGinn concludes that "it will be noumenal with respect to perception-based explanatory inferences" (*ibid*, 13-14). The term 'noumenal' is here used by McGinn in the more or less Kantian sense, namely, it alludes to the world beyond appearance, i.e., independent of our 'perception-based' apprehension of the world. McGinn concludes, therefore, that we are constitutionally unequipped to apprehend P, and as a result we are precluded from even constructing a psychophysical theory on the basis of inference to the best explanation. Consequently any theory of consciousness that must involve P is beyond our understanding. We are, as McGinn likes to put it, 'cognitively closed' to a theory of consciousness.

The conditions McGinn describes for the construction of a theory of consciousness seem to be beyond the capacity of even the conceivably most intelligent creature. If neither perception nor introspection can allow a creature to apprehend P, as required according to McGinn, the task appears to be impossible in principle. That is to say, a theory of consciousness is *absolutely* cognitively closed. Indeed, McGinn accepts that "if we suppose that *all* concept formation is tied to perception and introspection, however loosely, then no mind will be capable of understanding how it [consciousness] relates to its own body – the insolubility will be absolute" (*ibid*, 16). This possibility, I think, should trouble us. He claims that a naturalistic theory of consciousness is possible in principle. But how can we make sense of this claim if no creature could construct such a theory? We are left having to imagine a Fregean-type third realm where unthinkable theories exist in their own right. This fact makes his conclusion very implausible.

That said, McGinn speculates that it is 'just about' possible to conceive of some creature that is able to construct a naturalistic theory of consciousness independently of the faculties of perception and introspection (*ibid*). How? The creature would employ a highly developed faculty of *a priori* reasoning, the same faculty that we employ in a limited way with respect to numbers and other mathematical concepts, according to McGinn. Thus, he states that we can imagine this creature conceiving of consciousness, brains, and all the properties thereof independently of experience (*ibid*). This strikes me as a feat no less miraculous than the possibility of consciousness arising in brains causelessly that McGinn dismisses. Indeed, he surmises that the employment of such reasoning would likely be how God cognises (*ibid*). What this shows, more importantly, is that McGinn assumes that our intuitions, i.e., the basis of *a priori* reasoning, can

uncover how the natural world is. This assumption is representative of an attitude wholly discordant with that of naturalism.

1.4.1 Consciousness as a Non-spatial Property

Why does McGinn think that P, as the property of the brain that realises consciousness, is non-spatial? Many properties are non-spatial, e.g., the property of being worth ten dollars. But relational properties like this are only trivially non-spatial – there is no sense in which we think of them as spatial, i.e., they are not *space-implicating* – what precisely is meant by this phrase will be explained shortly. On the other hand, P is an intrinsic property of the brain, given that it is the property of the brain responsible for the brain's being conscious. McGinn tells us that P has to be non-spatial because consciousness itself is self-evidently non-spatial, and so as the property responsible for generating consciousness P must share this characteristic with consciousness. What is special about P, we must remember, is its epistemic role, namely, apprehending it allows us to link consciousness to the brain. And no spatial property can fulfil this role. McGinn explains that

> the senses are geared to representing a spatial world; they essentially present things in space with spatially defined properties. But it is precisely *such* properties that seem inherently incapable of resolving the mind-body problem: we cannot link consciousness to the brain in virtue of spatial properties of the brain (1991, 11).

But, why must we suppose that there is a quixotic property like P? Why can we not suppose instead that no such epistemologically handy property exists? Since the idea of an intrinsic property of the brain that is non-spatial is difficult to understand it seems plausible to deny that such a property exists. That would be unfortunate since it would seem to preclude us from being able to grasp directly the link between consciousness and the brain, but so be it.

McGinn's answer is that because consciousness is a natural, as opposed to supernatural, property it must be naturalisable in principle. He insists that "[t]here just *has* to be some explanation for how brains subserve minds... some theory must exist which accounts for the psychophysical correlation we observe" (*ibid*, 6). In other words, there must be some theory, appealing to laws of nature, by which how consciousness arises from the brain can be fully explained. And according to McGinn this theory concerns P. For this reason McGinn is confident that P exists, albeit beyond our ken. However, it should be emphasised that P plays a very specific epistemic role. McGinn's claim is that if we were able to grasp P (per impossible for McGinn) we would *immediately* understand how consciousness is related to the brain. By this act we would obtain complete epistemic satisfaction vis-à-vis an explanation of how consciousness arises from the brain. That is, we would gain a *feeling* of complete epistemic achievement in our understanding of consciousness as a natural property. But as Mark Rowlands notes, explanations do not always entail such epistemic satisfaction. Rowlands suggests that an explanation can be adequate without its necessarily being satisfying in this way, i.e., without its leading us to say such things as 'Now I've got it' or even 'Eureka!' (Rowlands 2001, 60). An explanation is said to be adequate, according to Rowlands' use

of the term, when it is rationally accepted as an explanation by us, that is, when we feel the explanation gives us a fuller understanding of the explanandum.

Essentially Rowlands argues that sometimes an explanation is adequate but not completely epistemically satisfying; that is to say, full-blooded epistemic satisfaction in the sense McGinn has in mind is not a necessary condition for an explanation to be adequate. Rowlands urges us to think of explanatory adequacy as ranging over a continuum, starting from full-blooded epistemic satisfaction passing to decreasingly satisfying explanations, until at some point the explanation is regarded as providing only a minimal understanding of the explanandum. He cites the example of explaining solidity, e.g., of metals, in terms of the ionic and covalent bonds between molecules (*ibid*, 62). This explanation is not epistemically satisfying insofar as it may not allow us to understand why, for example, ionic bonds are stronger than covalent ones. Moreover, even when this is explained, perhaps ultimately appealing to physical laws, further questions could still arise, that is, we might never attain epistemic satisfaction. He generalises the point as follows:

> The idea under consideration is that we can, in principle, render a particular explanation more epistemically satisfying than it is, or seems to be, by appeal to the various physical laws that underwrite the explanation. The problem is, however, that there is no *a priori* reason why physical laws should be any more epistemically satisfying than the explanations they underwrite (2001, 63).

The upshot of Rowlands' argument is that if an explanation in science does not have to be completely epistemically satisfying, then we do not have to suppose that an explanation of how consciousness arises from the brain involves P, namely, that property which would give us a completely epistemically satisfying explanation such that one could say 'Now I understand why the brain is conscious'. In other words, accepting that because consciousness is a natural property it must ultimately be explainable in natural terms, it does not follow that the theory used to explain consciousness must concern P as defined by McGinn.

1.5 A Difficulty with McGinn's Naturalism

It has already been suggested that the version of naturalism that McGinn espouses is incompatible with the one I have recommended. McGinn, like Chalmers, assumes that science aims to reconcile our intuitive understanding of the world with how the world is as we experience it. We intuitively understand consciousness as non-spatial and the brain as spatial, and the fact that science seems incapable of reconciling this conflict is seen by McGinn as a measure of the failure of science, that is, as testament to the limitations of our science. This is to think of our understandings in this sense as data, i.e., as independent facts in the world. Thus, McGinn thinks that the non-spatiality of consciousness is a given, something that science must accommodate in its theories. However, as I have argued, this approach is antithetical to that of science. Nothing of how we understand the world is given in this way. The intuitive plausibility of our understanding of a phenomenon, however strong, is simply *irrelevant* with respect to our scientific theories. Our intuitive understandings have no determinative role in the

confirmation of our theories. Again, we saw this with the example of calculus: how we intuitively understand the infinitesimal spatiotemporally cannot be reconciled with the theorems of calculus, but this fact has not stopped mathematicians from holding these theorems as true. Rather, they have worked to develop a conception of the infinitesimal that fits with the truth of the theorems. But McGinn contends that consciousness is self-evidently non-spatial, such that science *must* accommodate this fact. It is very difficult, however, to see how McGinn can hold both that consciousness is non-spatial and natural.

If one were to espouse some form of mind-body dualism one would have little problem in supposing that consciousness is non-spatial, as a non-physical property, but McGinn denies that he is a dualist. But it is not clear how intelligible this position is. The non-dualist must hold that there cannot be disembodied consciousness. But if being conscious depends on the physical brain in this way, then it is hard to understand how it is not spatial. Obviously we must be clear on what is meant by a property being spatial exactly. We do not ordinarily talk of properties as being spatial. But a plausible way of understanding what is meant by saying a property is spatial is to assert that it can only be attributed to an object said to be spatial. So, for example, blue is a spatial property because it cannot be attributed to a non-spatial object, e.g., the number two. Or conversely, we noted that the property of being worth ten dollars is non-spatial; and this fits with this criterion since, for example, something non-spatial like one hour of someone's labour can be said to be worth ten dollars. Anthony Brueckner and E. Alex Beroukhim offer a useful definition of a property's being spatial in this sense in terms of its being space-implicating; where "a property F is *space-implicating* iff necessarily, if x instantiates F, then x is spatial" (2003, 404). That is, a property is spatial only if it is

necessarily realised by a spatial object. But while McGinn holds that consciousness is realised, i.e., instantiated, by the brain, is it the case that he must suppose that it is necessarily realised by such a physical object? If he does not have to suppose this, then it seems plausible for him to assert that consciousness is indeed not space-implicating and *a fortiori* non-spatial. Clearly, to say that consciousness is not a space-implicating property in this sense is to assert that there are possible worlds in which consciousness is not realised by a physical object. Unfortunately, this suggests that disembodied consciousness is possible. Again, as Brueckner and Beroukhim point out, this is not a conclusion McGinn countenances (*ibid,* 398-99). McGinn recognises that the possibility of disembodied consciousness makes it difficult to understand how consciousness is related to the rest of nature, i.e., to other natural properties; thinking of consciousness as natural would be made impossible. It seems, therefore, that by asserting that consciousness is non-spatial McGinn's position collapses into the dualism he denies. But if he were to give up this claim, then his entire argument for his non-constructive naturalism fails.

Perhaps, in reply, one could argue that consciousness itself is not realised by the physical properties of the brain but rather by P. Hence consciousness does not have to be construed as a space-implicating property. But, of course, by holding that P is non-spatial the difficulty is merely transferred to P; that is, if one holds that P is not space-implicating then it could be instantiated non-physically, which implies in turn that consciousness can also be. Therefore, the possibility of disembodied consciousness is not averted.

But, perhaps McGinn only needs to assert that consciousness is non-spatial to the extent that it is imperceptible in principle. Thus, he might be willing to suppose that consciousness is space-implicating, but to note that just as we cannot perceive something as being worth ten dollars, so we cannot perceive something as being conscious. However, this comparison seems spurious. As noted earlier, the property of being worth ten dollars is trivially imperceptible, i.e., non-spatial, given that it is relational, whereas consciousness is non-trivially so because it is thought of as an intrinsic property of a physical object. Consciousness is realised in virtue of the physical constitution of the object in some crucial sense. Still, as McGinn points out, we cannot perceive the brain as conscious, rather than as soggy and grey, for example. But what is interesting about this observation is that the term 'consciousness' is not ordinarily applied to the brain; rather, we ordinarily think of people, and perhaps some other creatures, as being conscious. Indeed, we usually judge that S's brain is conscious only if S, its owner, is conscious. And to the extent that we judge people to be conscious we do so on the basis of *observations* about them. Therefore, consciousness, thought of as a property of people and other creatures, is observable. In this regard Nagel, for example, famously asserts that bats are conscious and that consequently there is something that it is like to be a bat.[17] If, however, consciousness is understood purely phenomenologically, that is, if we each only understood being conscious from the first-person viewpoint, then we could not assent to Nagel's claim. But we *can* assent to it, as Nagel does. The very question 'what is it like to be a bat?' would make no sense if consciousness were understood purely phenomenologically. A great deal more will be said on this issue in chapter 4.

---

[17] See Nagel 1974. Nagel's views are discussed in some detail in the chapter 2.

1.6 Summary

In this chapter I have outlined naturalism characterised by the treatment of philosophy as being continuous with science. This continuity follows from the fact that philosophical hypotheses, ultimately justified by our intuitions, cannot overrule our scientific hpotheses. In section 1.1 I illustrated this proscription using two historical examples of failed attempts by philosophers to undermine scientific or mathematical theories by appeal to our intuitive understandings of certain phenomena. Specifically, we looked at Descartes' denial of atomism based on his intuitive understanding of space and at Berkeley's attack on calculus based on its incompatibility with our intuitive understanding of continuity in spatiotemporal terms. Consequently, I argued, it behoves us to follow the precept: How we think of the world to be should ultimately be measured according to our experiences. And so how we think of the world should *not* be governed by our intuitions, however compelling these might be.

In section 1.2 I pointed out that this rule is often not followed with respect to one particular phenomenon in the world, namely, consciousness. Because our intuitive first-person understanding of consciousness is so compelling – it is how we are immediately acquainted with consciousness – there is the temptation to insist that our scientific theories must accommodate it. And given that consciousness understood in such phenomenological terms seems patently to resist naturalisation in this way, a scepticism prevails. This scepticism is usually expressed in terms of consciousness being unexplainable in physical terms. In this section we looked at one of the most notable sceptics in this regard, namely, Chalmers. In the case of Chalmers I showed how his non-naturalistic attitude leads him to make extravagant metaphysical claims that are

unfalsifiable. Indeed, I argued that his proposed theory of consciousness, based on this scepticism, is antithetical to science in the same way that Descartes' anti-atomism was.

In section 1.3 I pointed out that matters are confused by the fact that these sceptics call themselves naturalists. Both Chalmers and McGinn call themselves naturalists because they hold that everything is in principle naturalisable. However, they claim that science, at least as it is practised, cannot explain consciousness. According to Chalmers consciousness can be naturalised only if how we practise science is suitably reformed; and this requires our conceiving of consciousness as a fundamental property. This reform, I argued, requires us absurdly to hold as true theories that are empirically untestable. According to McGinn, while consciousness is in principle naturalisable our innate cognitive limitations rule out our ever being able to construct a theory of consciousness. Underpinning the claims of both these philosophers is the assumption that science must accommodate our intuitive understandings of consciousness as determined from the first-person viewpoint. And in this sense their views are not naturalistic, i.e., they do not adopt a naturalistic *attitude*.

In section 1.4 I looked at the plausibility of McGinn's view. I noted his claim that consciousness is likely naturalisable only for a creature that can construct a theory of consciousness *a priori*. This claim, I pointed out, betrays a staunchly non-naturalistic attitude. Further, we saw the consequences of this attitude in terms of his central claim that consciousness is self-evidently non-spatial – here he appeals to our intuitions. If consciousness is indeed taken to be non-spatial we concluded that we cannot rule out the possibility of disembodied consciousness. This possibility, however, is incompatible with

a naturalistic worldview, since consciousness would have to be in some sense independent of other natural phenomena.

Chapter 2

The Problem of Consciousness

In the last chapter we looked at subsuming a phenomenon under our scientific theories, that is, naturalising a phenomenon. There I argued that our intuitions cannot determine our naturalistic conceptions of phenomena. If our conception of a natural phenomenon according to a scientific theory is contrary to our intuitive understanding of it, so much the worse for our intuitive understanding. With respect to consciousness I noted that our intuitive understanding of it derives from our immediate acquaintance with it from the first-person viewpoint. This phenomenological understanding of consciousness does not seem to square with how we think of the world in physical terms. This essentially describes the problem of consciousness – there seems to be no way of understanding consciousness as a physical property, and hence we cannot rule out the possibility that it is not physical.

Now, given the remarks above the obvious reply is that ultimately our intuitive understanding of consciousness is irrelevant and so there is no problem. What complicates matters, however, is that, as noted in the last chapter, our intuitive first-person understanding of consciousness seems ineliminable. While, as we saw vis-a-vis

calculus, we can think of continuity in analytical terms rather than spatiotemporally, the same approach to consciousness appears impossible. We seem forced to accommodate our intuitive phenomenological understanding of consciousness in our naturalisation of it. My aim in this chapter is to face up to the strength of this intuitive understanding of consciousness. It is this understanding that underpins the problem of consciousness.

The purpose of this chapter, then, is to evaluate the problem of consciousness as a barrier to the possibility of naturalising consciousness. The problem, as was noted in the introduction, is usually delineated in terms of being a problem for physicalism. As such it has been used by some to argue that physicalism is either false or open to serious doubt. Below I shall look at the most notable anti-physicalist arguments in turn, namely, those offered by Thomas Nagel, Frank Jackson and David Chalmers. I argue that all three arguments fail to refute physicalism, which at least suggests that there are no *a priori* reasons for denying that consciousness is physical.

Here, I assume that the physicality of consciousness is a sufficient condition for its being naturalisable.[18] Naturalising a phenomenon is best understood as explaining it in terms of, or specifying its relations to, other natural properties. William Seager offers a useful definition. He states that a phenomenon, X, is naturalised if and only if (1) X has been explained in terms of something else, (2) the something else does not logically involve X, and (3) the something else is properly natural (Seager 1999, 249). Moreover, to naturalise a phenomenon is thought of as explaining it in *physical* terms. The underlying physicalist assumption here is eloquently expressed by Quine: "Nothing happens in the world, not the flutter of an eyelid, not the flicker of a thought, without

[18] I borrow this phrasing from Uriah Kriegel (see Kriegel 2005, 23, n. 1).

some redistribution of microphysical states" (1978, 98). A microphysical state is any state that can be described in terms of the fundamental entities posited by physics. This is to embrace physicalism roughly characterised by the claim that all the facts are detemined by the physical facts. What this amounts to exactly is a matter of dispute. However, perhaps the least contentious construal of physicalism vis-a-vis the mental is to assume the mental supervenes on the physical, as intimated by Quine's quote above. Here it is supposed that mental properties depend in some minimal sense on physical properties. This dependence is often understood in terms of covariation, namely, that any change in mental properties of some creature, say, necessarily implies a change in some of its physical properties.[19]

I shall look at the anti-physicalism of Nagel, Jackson and Chalmers in turn. Nagel, one of the first people to enunciate the problem of consciousness, is beholden to a full-blooded realism about the phenomenological features of experience. Consequently, he thinks that only by explaining how these features are physical properties can we assume physicalism is true. Below, I argue that his realism is too strong. It is based on the assumption that it makes sense to talk about understanding what it is like to be, i.e., to have the phenomenology of, some other kind of creature – his example is being a bat. I argue that we cannot in fact make sense of this idea, thus undermining the intuition he appeals to vis-a-vis the problem of consciousness.

Jackson likewise subscribes to a full-blooded realism about the phenomenological features of experience. He argues that our having conscious experiences furnishes us with direct knowledge of the phenomenological features of these experiences, but that this knowledge cannot be captured in physical terms. No amount of physical information

---

[19] See Kim 1994, 575-583.

about the world will tell us anything about the phenomenological features of experience. Therefore, he concludes, these features cannot be physical. In reply, following similar objections raised by Paul Churchland, I argue that the kind of direct knowledge or acquaintance we have of these features does not concern the kind of factual knowledge we have of the physical world. Consequently, he does not succeed in showing that the phenomenological features of experience, i.e., qualia, are non-physical properties.

Finally, Chalmers appeals to our intuitive understanding of consciousness to argue that we can always conceive of consciousness independently of the physical facts. Essentially, he presents an augmented version of the classic conceivability argument to the effect that because consciousness can always be thought of as independent of any physical properties it cannot be identical to any such properties. In other words, Chalmers argues that consciousness does not supervene on physical properties in the sense needed to supposrt physicalism. Instead, he argues that consciousness is only contingently dependent on the physical; that is to say, while in the actual world conscious states depend on physical states (in terms of what he calls 'informational invariance') it does not follow that in some world physically identical with this one consciousness even exists. This conclusion leads Chalmers to suggest that we can construct a theory of consciousness by thinking of conscious properties as fundamental in the sense of their existence being taken as a brute fact about the actual world. Such a theory of consciousness requires a revision of our sciences as mentioned in the last chapter. My response to Chalmers' view is two-fold. First, I rehearse some of the arguments against his conceivability argument, focusing on one given by Peter Carruthers. I support the conclusion that Chalmers' conceivability argument fails. Second, I point out that the idea

of thinking of conscious properties as basic is incoherent. Therefore, his proposed theory of consciousness, I argue, cannot work. Together the refutation of these three anti-physicalist arguments shows that despite appeal to our intuitive understanding of consciousness it has not been demonstrated that physicalism is false. In this respect, the possibility of naturalising consciousness is not ruled out.

2.1 Nagelian Anti-Physicalism

It is Thomas Nagel who first formulated the problem of consciousness in his classic article 'What Is It Like to Be a Bat?' (1974). He claims that, at the time of his writing this article, identity theorists had failed to consider seriously the subjective character of conscious experiences. For a creature to have conscious experiences, Nagel points out, there must be something that it is like to be that creature. This is the felt quality of experience. But to say that each conscious experience is identical with some physical, e.g., neural, state tells us nothing about its felt quality. This in itself is perhaps a trivial observation. But Nagel proceeds to offer reasons why we should think that, quite generally, physical descriptions of experience *cannot* tell us anything about their subjective character. Consequently we might at least suspect that the subjective character of experience is not physical.

He begins by characterising the subjective as that pertaining to a particular point of view, i.e., a first-person point of view. According to Nagel the subjective aspect, or character, of experience is only apprehensible from the first-person point of view. By contrast, Nagel notes, most properties can be apprehended from the third-person point of view; that is to say, grasping them is not dependent on any particular point of view. For

example, while a creature with a different perceptual apparatus may be unable to *see* that a table, say, is a certain height, it can nonetheless apprehend this feature of the table in other ways. Hence, the height of a table, as a physical property, is independent of any particular way of apprehending it; and as such this feature is independent of any particular point of view inasmuch as one's point of view is determined by how one perceives things in the world, as Nagel suggests. Accordingly, insofar as pain is identical with certain physical features, e.g., certain neurological properties, it is understood as being independent of any particular point of view. But, viewed phenomenologically a pain is defined in terms of its felt quality, and this feature seems to be essentially connected to the single point of view of the person having the sensation (1974, 437). Yet, so described the subjective character of experience seems to be essentially private, such that it must be seen as a property of experience that could never be known by another person. Hence, we slide into solipsism where I, at least, enjoy such felt qualities of experience, but there is no way for me to know if others do. Nagel resists such solipsism by insisting that there is a matter of fact about what it is like for either you or me to have conscious experiences. He states:

> I am not adverting...to the alleged privacy of experience to its possessor. The point of view in question is not one accessible only to a single individual. Rather it is a *type*. It is often possible to take up a point of view other than one's own, so the comprehension of such facts is not limited to one's own case. There is a sense in which phenomenological facts are perfectly objective: one person can know or say of another what the quality of the other's experience is (*ibid*, 442).

Nagel's claim is that the subjective characters of our individual experiences are alike. So, for example, the felt quality of person *A*'s toothache is like that of person *B*'s toothache. Why should we think this when such felt qualities are not apprehensible by others? Nagel assumes their felt qualities are alike because *A* and *B* are of the same species, based on their being physiologically alike in relevant ways. Therefore, as members of the same species we can know what each other's experiences are like, since we share the same type of first-person point of view.

But what of a different type of creature? Nagel assumes that mammals and perhaps other higher animals are conscious, and therefore there must be something that it is like to be each of these creatures. Can we know the subjective characters of their experiences? He considers the case of a bat. As a mammal it is plausible to assume that there is something that it is like to be a bat, that is, to have bat experiences. However, certain of their experiences are quite alien to our own. Specifically, he is thinking of bat sonar. Since we do not have anything like this kind of perception it is difficult for us to imagine what it is like to echolocate. Imagining what it is like to be a bat, as Nagel describes it, involves extrapolating "to the inner life of the bat from our own case" (1974, 438). Thus, to the extent that our experiences are different from those of bats, we cannot extrapolate from our own case. We cannot imagine what it is like to be a bat, in this sense, by imagining *ourselves* as bats, e.g., our hanging upside down, our flying about. What is demanded is our being able to conceive what it is like from the bat's point of view to do these sorts of things, specifically to echolocate.

This difficulty can be generalised. Given that phenomenological facts are only apprehensible from a particular point of view, they are constitutively dependent on, i.e., essentially connected to, the type of point of view from which they are apprehended. They are, after all, facts about having a specific type of point of view. Again, in contrast, physical facts are conceived objectively, that is, independently of any particular point of view. Clearly, therefore, if phenomenological facts can only be thought of as concerning a particular point of view, there is little hope of conceiving them objectively. This simply seems to be ruled out. Indeed, with respect to subjective phenomena in general, Nagel observes that "any shift to greater objectivity – that is, less attachment to a specific viewpoint – does not take us nearer to the real nature of the phenomenon: it takes us farther away from it" (*ibid*, 445).

We have already noted the contrasting nature between phenomenological and physical concepts in this regard. Nagel gives the example of lightning. While the apprehension of lightning requires a first-person point of view, it is not understood according to how it is perceived. So if we were to imagine a Martian that has different perceptual capacities from us, e.g., perhaps it perceives only ultraviolet light, it is still possible to assume that it could understand the concept of lightning. Lightning is not essentially understood as a bluish flash in the sky, rather this is merely one of its modes of presentation (*ibid*, 443). It is in this sense that a physical concept like lightning is independent of any particular point of view. Thus, as Nagel explains, "it seems inevitable that an objective, physical theory will abandon that [first-person] point of view" (*ibid*, 437). That said, the identity theorist maintains that conscious experiences are physical states. If this is so, Nagel argues, then their subjective characters should be reducible to

physical properties, i.e., we should be able to explain them objectively in physical terms. But, as we have seen, there seems to be no possibility of doing this.

According to Nagel this contrast between phenomenological and physical concepts presents a difficulty for physicalism. If we wish to hold physicalism true, then we need to explain how the subjective characters of experiences quite generally are related to physical properties. Accordingly, we need to explain the subjective characters of a bat's echolocatory perceptions, for example, in physical terms; that is to say, we should provide a physical, i.e., objective account of what it is like to have such bat experiences *for a bat*. However, a physical account of the subjective characters of such experiences requires abandoning the particular point of view of the creature in question; and since these characters are dependent on this point of view it does not seem possible to provide such an account. Therefore, inasmuch as this is not possible, there is no reason to hold that physicalism is true. Nagel concludes that the onus is on the physicalist, e.g., the identity theorist, to demonstrate how a physical account of the subjective character of experience is possible, that is, "[i]f physicalism is to be defended, the phenomenological features must themselves be given a physical account" (*ibid*). He therefore expresses the kind of mysterianism to which Joseph Levine later subscribes, namely, that as it stands we are not in a position to know how physicalism might be true.[20] Nagel, in his later writing, tends to the conclusion that physicalism is a misguided doctrine, for example, when he states:

> It is the phenomena of consciousness themselves that pose the clearest challenge to the idea that physical objectivity gives the general form of reality. In response I want

> not to abandon the idea of objectivity entirely but rather to suggest that the physical is not its only possible interpretation (1986, 16-17).

But in the earlier article under discussion he tempers his anti-physicalism by suggesting that our inability to explain consciousness in physical terms might result from conceptual limitations. He considers the claim that matter is energy. Unlike current physicists, rather than the lay person, the pre-Socratic philosopher did not have the concepts to enable him to understand this claim. Nagel suggests that with respect to the claim that consciousness is physical we might be in the same position as the pre-Socratic philosopher with respect to the claim that matter is energy. But perhaps in the distant future people will have developed the concepts that will allow them to see how consciousness is physical. For now we remain in the dark.

Nagel speculates on how one might provide such a physicalist account. What is demanded, he suggests, is a way of conceiving of such phenomenological features objectively, i.e., independently of any particular point of view, without at the same time abandoning the particular point of view that constitutes it. This appears to be plainly contradictory. It describes a necessarily impossible task. However, rather than reaching this conclusion Nagel argues that it *could* be that this task merely seems impossible to us because of our current conceptual framework. We cannot rule out the possibility that there is a way of understanding the subjective character of any creature's experience objectively, but that either we presently lack the concepts needed to gain such an understanding, or we are perhaps cognitively closed to these concepts (*ibid*, 440). In this respect Nagel embraces an agnostic mysterianism.

---

[20] See Levine 1983.

Despite the explicit denial that his argument shows that physicalism is false it is often interpreted as effectively showing this. As noted, the paradoxical nature of providing an objective account of the subjective character of experience seems to make defending physicalism impossible, rather than difficult. To the extent that we must think of consciousness as being dependent on a particular point of view we *cannot* suppose physicalism is true if physicalism demands giving up that point of view. Indeed, in later writings Nagel takes this view. He diagnoses physicalism as suffering from what he calls "objective blindness", that is, it presupposes that all facts can be understood independently of a particular point of view (1986, 7).

Should we conclude, therefore, that physicalism is false? Again, the provision of a physicalist account of these phenomenological features seems plainly paradoxical, in the sense of being self-contradictory. The development of concepts that bridge the first- and third-person viewpoints seems hopeless in a way that our speculations concerning other conceptual advances do not. For example, we presently do not have the conceptual wherewithal to explain how gravity relates to electromagnetism and the weak and strong nuclear forces, nonetheless there is nothing paradoxical about the development of such concepts.

Crucially, the paradoxicality of understanding the subjective character of experience in physical terms points to another possibility, namely, that the very idea of understanding what it is like to be another creature is incoherent. Indeed, below I argue that it is incoherent. Moreover, its incoherence suggests that Nagel's doubts about physicalism are unfounded. The worry is that the subjective character of experience depends on a particular point of view implying that it is impossible to understand, and *a*

fortiori to know, facts concerning such phenomenological features except from the first-person point of view. But this is impossible, I contend, because there is no sense in which we can know these facts. That is to say, rather than our being unable to gain such knowledge there is nothing for us to know or understand vis-a-vis facts concerning what it is like to be some creature.[21]

Nagel thinks that it is possible in principle to know what it is like to be an alien creature given that some being unlike ourselves could imagine it. He writes:

> From the perspective of one type of being, the subjective features of the mental states of a very different type of being are not accessible either through subjective imagination or through the kind of objective representation that captures the physical world... A being of total imaginative flexibility could project himself directly into every possible subjective point of view, and would not need such an objective method to think about the full range of possible inner lives (1986, 17).

But what grounds does Nagel have for assuming that such a being could exist? He plausibly defines imagining in this sense as extrapolating to the inner life of another type of creature from one's own case. If this is impossible for us to do in the case of bats because we are so different from them, how are we to suppose that this being – itself necessarily different from some other creatures – is able to do this instead? What faculty

---

[21] A similar line of argument is taken by Yujin Nagasawa. Nagasawa points out that if it is impossible in principle to know what it is like for a bat to be a bat, then physicalism is not threatened. Here it is assumed that physicalism might be false because there are facts about bats that the physicalist cannot know, i.e., facts concerning the phenomenological

does this being possess that we do not? Nagel has nothing to say about this. This being's imagining what it is like to be another creature rests on *extrapolating* from its own point of view, and this is a limitation – relative to its own point of view some creatures are always going to be sufficiently unlike it to make such extrapolations less successful. Having a point of view puts limits on one's imaginative ability no matter what kind of being you are. Likely, by the phrase 'total imaginative flexibility' Nagel is attempting to describe a capacity to imagine *beyond* the limitations of a particular point of view. But this capacity amounts to the ability to apprehend another point of view directly, that is, without having to extrapolate from one's own. Such a capacity can only be thought of as divine since it entails powers unavailable to finite beings.

Let us suppose that the being Nagel presumes is ideal, and as such it is in essence divine. It seems possible by this measure to think that there may be facts such as what it is like to be another kind of creature that are knowable by such a being, granting these facts are unknowable by us. In similar vein we might suppose that any infinite decimal expansion of a transcendental number, such as π, is determinate such that it is at least graspable by God or some infinite being. This is indeed what Nagel supposes while of course, as finite beings, we cannot grasp such an infinite expansion.[22] By this measure for example, whether or not the sequence '7777' occurs in the infinite decimal expansion of π is in principle determinable. However, we cannot determine this not because of the limited calculating ability of our brains, but rather because we cannot make sense of the idea of this sequence occurring inside an *infinite* decimal expansion – an expansion

---

features of their experiences. However, these facts cannot be known whether physicalism is true or not, since it is impossible to know them in principle (see Nagasawa 2004).

[22] This example, as presented by Wittgenstein, is discussed by Nagel. See Nagel 1986, 107.

which we must think of as limitless. As Wittgenstein points out, there is no inside an entity without limits.[23] Perhaps God has the *imaginative ability* to think of this sequence occurring inside the infinite decimal expansion, but we can never conceive of this; and because we cannot do so, the question 'Does "7777" occur inside the infinite decimal expansion of π?' is nonsensical and hence unanswerable.

Likewise, because we cannot even conceive of imagining what it is like *for a bat* to be a bat, for example, the question 'What is it like for a bat to be a bat?' is nonsensical. Our inability in this respect does not concern our limited cognitive capacity, rather it concerns the conceptual impossibility of any non-bat creature *imagining* what it is like for a bat to be a bat in the way Nagel assumes to be possible. The impossibility of this task is illustrated by an example originally given by Bernard Williams. First, Williams thinks of his imagining himself *as* Napoleon; so like an actor on stage he plays the role of Napoleon, imagining himself as a French general, perhaps surveying the destruction at Austerlitz or spending his days in exile on the island of St Helena. Imagining such things seems possible. Then, however, he tries to think of himself *being* Napoleon, that is, his having the same physiology, history, character, etc. of the actual Napoleon. Williams puzzles: "What could be the difference between the actual Napoleon and the imagined me?" (1973, 42). His point is that insofar as there is no discernible difference we must doubt that he can imagine *himself* being Napoleon. We cannot make sense of this idea, it is incoherent.

There is no sense in which we can come to know facts about the subjective character of the experiences of other creatures, e.g., a bat. Our not being able to make sense of knowing such facts must be distinguished from our being unable to know them. Its

---

[23] See Ambrose, 195-199.

nonsensicality in this sense points to there being nothing to know or understand. Now, one might object that intuitively we think of such facts as knowable. We feel there is a matter of fact about what it is like to be a bat. But, again, there is no sense in which we can understand these facts as such, just as we cannot understand there being a fact about whether '7777' occurs in the infinite decimal expansion of π or not. What is it like for a bat to be a bat? Again, the question is unanswerable. Is this to deny that there is something that it is like to be a bat? To paraphrase Wittgenstein, there is not something that it is like to be such a thing, but that is not to say there is nothing that it is like to be one.[24] What it is like to be a bat is not *something* in the sense of being identifiable by others. There is, nevertheless, something that it is like to be a bat in the most general sense that there is a world for a bat, rather than nothing at all.

To sum up, Nagel points to a crucial characteristic of conscious experience that his predecessors and peers seemed to overlook, namely, that the phenomenological features of experiences are dependent on a particular point of view. Nagel worries that this characteristic of experience makes it difficult for us to understand and therefore come to know facts concerning these features. I have argued above that the paradoxical nature of understanding from a third-person point of view these phenomenological features that depend on a particular point of view should be taken seriously. It suggests that asking what it is like to be such-and-such a creature is nonsensical – such questions are unanswerable. Seeing this endeavour as nonsensical rather than difficult to answer undermines the threat to physicalism that Nagel worries about. Physicalism cannot be false insofar as it cannot explain facts concerning the phenomenological features of

[24] See Wittgenstein 1958, 304.

experience, as Nagel reasons, because there is no sense in which these facts can be explained.

2.2 Jackson's Knowledge Argument

We have seen that Nagel presents an epistemological problem for physicalism that he believes threatens it. By contrast Frank Jackson argues outright that physicalism is false. Although Jackson is concerned with the difficulty addressed by Nagel, he sees the focus of his argument to be quite different. He takes Nagel to be concerned with the problem of describing what it is like *to be* a particular type of creature in physical terms (see Jackson 1982, 131-32). Jackson, on the other hand, points to the problem of explaining the subjective character of any *particular* experience in physical terms, something that he thinks must be possible if this feature is physical. His argument can be summed up as follows: No amount of physical information can capture what it is like for a person to have a certain type of experience, e.g., the experience of redness looking at a ripe tomato. So facts about our experiences concerning these kinds of phenomenological features cannot be known in physical terms. Therefore, these features are not physical in nature. Hence, physicalism is false. Below I argue that Jackson's argument fails because it rests on an invalid inference he makes concerning acquaintance with the phenomenological qualities of our own experiences and knowledge of the phenomenological qualities of other people's experiences. This problem is made clear when we consider Paul Churchland's closely related objections to Jackson's argument.

Jackson asks us to imagine an individual he calls 'Fred', who reports being able to distinguish two colours with respect to the wavelengths of light we describe as the red

spectrum. His claim is supported by evidence, namely, he is able repeatedly to separate a pile of ripe tomatoes into two consistent groups; where these tomatoes appear to everyone else as being a single colour. As well, careful study of his physiology concerning vision reveals he is abnormal in relevant ways. We might assume, therefore, that Fred realises at least one distinctive colour quale with respect to what he calls 'red_1' and 'red_2', that is, the distinct colours that we only see as red. Jackson notes that no amount of physical information about the differences between Fred and the rest of us describes how red_1 and red_2 feel to him (*ibid*, 129). Physical information is any information used to describe functional roles of, in this case, Fred's visual processes in relation to physical objects in the world. This is information about the world we gain from the core natural sciences of physics, chemistry, and biology. There is something about Fred, i.e., a property of some of his colour experiences, that no amount of physical information can describe.

One might perhaps doubt that this putative property of some of Fred's colour experiences really exists, since it cannot be described in functional or physical terms. But, in reply, Jackson imagines someone else acquiring the capacity to see the new colour, or colours, by surgical means; so that after the operation she declares "So that is what it is like to see the colours that Fred saw." This suggests that she learns something genuinely new about Fred. In the same vein Jackson offers another example concerning a neuroscientist called 'Mary', who is raised from birth in a black-and-white environment. Mary has learned everything there is to know about the physical processes concerning vision; that is to say, she knows *all* the physical information there is about vision. Yet, if she is suddenly presented with a colour, there is no doubt, according to Jackson, that she

would learn something new, namely, what it is like to see this colour, i.e., some colour quale. And given that she knows everything physical about vision, this new thing that she would learn cannot be a physical property. Hence, physicalism is false.

Most of the objections to Jackson's argument centre on the example of Mary. So this is where I shall also concentrate my efforts. Consider one of the objections leveled by Paul Churchland. Churchland offers the following construal of Jackson's argument (1985, 23):

> (1) Mary knows everything there is to know about brain states and their properties.
>
> (2) It is not the case that Mary knows everything there is to know about sensations and their properties.
>
> Therefore, by Leibniz's law,
>
> (3) Sensations and their properties $\neq$ brain states and their properties.

Churchland points out that (3) only follows from (1) and (2) so long as the phrase "Mary knows" is employed univocally in these two premisses. But this is not the case. Premiss (1) concerns propositional knowledge while (2) concerns knowledge by acquaintance. Given this fact it is entirely possible that the properties of Mary's brain and those pertaining to her colour experiences are one and the same. Churchland determines that at most we can say that Mary, before seeing the colour red for the first time, "does *not* have a representation of redness in her prelinguistic medium of representation for sensory variables" (1985, 24). In other words, she is not acquainted with the sensation of redness *for herself.* And when premiss (2) is expressed in this way, without using the term 'know

about', we can see that (3) does not follow from it in conjunction with (1). Therefore, Jackson's argument plausibly interpreted in this way is invalid.

In response to Churchland's objection above (Jackson 1986) Jackson says that Churchland's construal of his knowledge argument is not accurate. Jackson agrees that Mary learns a new fact about herself when seeing something red in the trivial sense that she has never had this kind of experience before. He observes that before being released from her black and white environment Mary "could not have known facts about her experience of red, for there were no such facts to know" (1986, 130). What is crucial, however, is that upon seeing the colour red for the first time Mary gains knowledge about the property of her own experience, which is likewise a property of *other* people's experiences according to Jackson. Since the physicalist assumes that Mary already knew about all the physical properties of people's experiences, it should not be possible for Mary to learn anything new in this respect. The fact that Mary would do so indicates that this property that Mary learns about cannot be physical.[25] What matters, Jackson thinks, is that after seeing the colour red for the first time Mary realises that she did not know (howsoever) about certain properties of other people's visual experiences, i.e., red qualia. Accordingly, he offers a more accurate formulation of the knowledge argument (1986, 293):

> (a) Mary (before her release) knows everything physical there is to know about other people.

---

[25] This assumption is challenged by Daniel Dennett, who argues that if indeed Mary knows everything about hers and other people's visual experiences in physical terms, then she would not learn anything new when seeing the colour red for the first time. Jackson assumes that she could not know about what he calls red qualia, but he provides no independent argument for this claim. He simply appeals to our intuitions in this respect (see Dennett 1991, 398-406).

(b) Mary (before her release) does not know everything there is to know about other people (because she *learns* something about them on her release).

Therefore,

(c) There are truths about other people (and herself) which escape the physicalist story.

The suggestion is that by construing the argument in this way the charge of equivocation is deflected since we can see, as Jackson explains, that "the point is not the kind, manner, or type of knowledge Mary has but what she knows" (*ibid*).

But, as Churchland notes in reply to Jackson's defence of his argument, this more accurate formulation of it still does not escape the charge of equivocation (see Churchland 1998, 143-45). He notes that (a) is an implicit conditional that says that 'if there is any property of people and this property is physical, then Mary knows about this property'. In (b) we are told that there is one thing that is a property of people that Mary does not know about. That is to say, the consequent of the conditional implied by (a) is false. Accordingly, we infer that its antecedent is false, that is, it is not the case that all of the properties of people are physical, hence (c). Churchland points out, however, that inferring (c) from (a) and (b) is only legitimate when the phrase 'know about' is used univocally in (a) and (b). And, once again, this is not the case. Therefore, Jackson's formulation of the knowledge argument is invalid[26].

---

[26] Churchland insists that the burden is on Jackson to provide a version of his argument that avoids this equivocation. Churchland himself suggests one way of trying to do this, namely, by quantifying over all the possible ways of knowing about properties. The suggestion involves claiming that if there is any property and it is physical, then Mary knows about it in each and every way possible. While equivocation is thereby avoided the universally quantified conditional expressed by this claim is too strong. Mary, after all, lacks one way of knowing about the properties of people because of her upbringing (see Churchland 1998, 149-51 and 153-57).

The equivocation that Churchland points out in Jackson's improved formulation of the knowledge argument is clear enough. Indeed, Churchland is prompted to remark: "Here I am surprised that Jackson sees any progress at all with the... formulation, since I continue to see the same equivocation found in my earlier casting of his argument" (1998, 144). Why would Jackson not have anticipated this difficulty given that this formulation is meant to defend against the charge? I think the answer lies with premiss (b). There he stresses that after her release Mary learns something new about *other* people. Crucially, he not only assumes that Mary comes to know by acquaintance what it is like for her to see the colour red, he also explicitly claims that she gains *factual* knowledge about other people's visual experiences. And this factual knowledge is propositional in kind. He supposes that once Mary has a sensation of redness she is justified in asserting that thereby she knows that other people's experience of seeing the colour red has 'this property', i.e., the red quale she is said to realise at the moment in question. Before she lacked this propositional knowledge, that is, she did not know *that* other people's red experiences have this property. Thus, since the knowledge she lacked before her release is propositional, the phrase 'know about' in (a) and (b) is used univocally.

But how exactly is Mary *justified* in believing that other people's colour experiences have this property, i.e., as she is acquainted with it, given that her evidence is limited to her own case? Jackson equates doubting that this property is realised by others to scepticism about other minds. Accordingly, it is as much a mistake for Mary to doubt that by experiencing redness she learns something new about other people's colour experience as it is for her to doubt that other people have minds (1986, 294). But here we need to distinguish between her knowing that other people's experiences of red have *this* i.e., a

phenomenological quality understood indexically, and her knowing that they have some phenomenological quality. Scepticism about other minds concerns this latter kind of knowledge, i.e., doubting that other people's experiences of red have *any* phenomenological quality at all. If, on the other hand, we grant that other people's experiences of red have some phenomenological quality it does not follow that *this*, i.e., the quale with which Mary is newly acquainted, is the same type that others enjoy. Appeal to the implausibility of scepticism about other minds, therefore, does not permit Mary to assume that she knows that other people's experience of red have the same phenomenological quality as her own.

More generally, we noted that Jackson assumes that this knowledge that Mary gains is propositional in kind. He supposes that Mary comes to know that other people's experiences of red are like *this*. However, knowledge about *p* is propositional to the extent that it is justifiable, that is, to the extent that it is possible in principle to provide reasons for believing that *p*. Mary cannot provide any such reasons with respect to her supposedly new knowledge about *other people's* experiences. That is because her knowledge of the phenomenological quality of *her own* experience of red is noninferential in nature, that is, she comes to have it directly. Therefore, she has no means by which to justify her knowing about other people's experiences. At best, as Jackson suggests, she might appeal to our not doubting seriously that other people have minds. But as we have seen, such an appeal is not justification for the knowledge claim Mary is imagined to make.

To sum up, despite her rejection of scepticism about other minds Mary can still plausibly doubt that the red quale she realises is the same as that of other normally

sighted people. This fact tells us that Mary is not justified in assuming that she knows *that* others realise the same kind of red quale as she does. In other words, by her new colour experience she does not gain new propositional knowledge about other people, as Jackson claims. Therefore, his claim that what Mary learns on seeing colour for the first time is a fact about other people is spurious. More generally, the charge of equivocation still holds, so that Jackson's knowledge argument fails to refute physicalism.

2.3 Chalmers' Panpsychism

We have seen that neither Nagel's scepticism about physicalism nor Jackson's anti-physicalist argument are successful. But, more recently David Chalmers has offered some interesting arguments against physicalism which have become influential. It therefore behoves us to consider Chalmers' views in some detail.

How we think of consciousness seems utterly distinct from how we think of the physical world. Nothing about how we understand consciousness, it seems, falls under a strictly physical description. While, in response to this fact, it may be plausible to think of consciousness as supernatural, that is, as something that transcends the physical world, such thinking is wholly counter to our scientific worldview. The success of the sciences quite generally suggests overwhelmingly that we have no need to appeal to supernatural forces to explain phenomena. Perhaps the death knell for supernaturalism in this regard was Charles Darwin's explanation of the origin of species in wholly naturalistic terms. Thus, despite its seeming incommensurability with physical phenomena we endeavour to naturalise consciousness, that is, to bring it under our sciences.

But this incommensurability points to an underlying difficulty for the project of naturalising consciousness. As David Chalmers notes: "All sorts of mental phenomena have yielded to scientific investigation, but consciousness has stubbornly resisted. Many have tried to explain it, but the explanations always seem to fall short of the target" (1995, 200). This constitutes what Chalmers dubs the 'hard problem' of consciousness. He diagnoses the problem in very broad terms as follows:

> It is undeniable that some organisms are subjects of [conscious] experience. But the question of how it is that these systems are subjects of experience is perplexing. Why is it that when our cognitive systems engage in visual and auditory information-processing, we have visual or auditory experience: the quality of deep blue, the sensation of middle C? How can we explain why there is something it is like to entertain a mental image, or to experience an emotion? It is widely agreed that experience arises from a physical basis, but we have no good explanation of why or how it so arises. Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does (*ibid*, 201).

The problem, as Chalmers sees it, originates from the nature of explanations used in the natural sciences. The natural sciences standardly employ reductive explanations which involve defining phenomena functionally. More precisely, the scientist explains some phenomenon, X, by specifying the mechanism that realises X's functional role. For example, reproduction can be analysed in terms of its functional role, i.e., roughly to the ability of an organism to produce another organism (see Chalmers 1996, 43-44).

Accordingly, a fleshed out analysis of this sort constitutes an explanation of the phenomenon. However, such reductive explanations fail with respect to conscious experience. Chalmers contends that any detailed descriptions of the mechanisms in the brain involved in cognitive processes will not explain why these mechanisms give rise to conscious experience. Thus it makes sense to ask: "Why doesn't this information-processing go on 'in the dark', free of any inner feel?" (*ibid*, 203). Nothing in descriptions of such mechanisms seems to rule out this possibility.

Chalmers talks of conscious experience as an additional phenomenon, as something that arises from physical processes but which does not reduce to them. He thinks of conscious experiences as over and above the brain states said to realise them; as when he describes them as *accompanying* the performance of certain cognitive functions (*ibid*). But given that the rest of nature falls under physical descriptions, the obvious way to naturalise consciousness is to assume it too is a physical phenomenon. This requires us to suppose that consciousness is identical with certain brain states, or perhaps identical with the functional roles realised by such brain states. Accordingly when Chalmers notes that there is wide agreement that experience arises from a physical basis, but that we cannot explain how this is so, we might reply that the reason every experience 'arises' from some physical state is that it is identical with that physical state, or with the functional role this physical state occupies.

Chalmers' main reply to this physicalist identity thesis is to argue that any identity between conscious states and brain states is always conceivably false, *and therefore their distinctness is always a logical possibility*; or as he would prefer to say conscious states do not supervene logically on physical states (see Chalmers 1996, 93-122). Briefly, to say

that *B*-properties supervene *logically* on *A*-properties is to assert that given an account of the facts concerning *A*-properties we can know *a priori* all the facts concerning *B*-properties (*ibid*, 35-36). Moreover, following Saul Kripke, Chalmers notes that identity is a necessary relation, i.e., something has to be identical with itself in *all* possible worlds (see Kripke 1980). By this measure, it must be *logically* impossible for a phenomenal property, i.e., a quale, to be non-physical. However, according to Chalmers, it is plain that we can conceive of possible worlds in which everything is physically identical with the actual world, but where nothing is conscious, i.e., a zombie world. Therefore, such psychophysical identities cannot hold, hence physicalism is false. Thus, he claims that consciousness is ontologically distinct from the physical; in other words, he countenances dualism. However, as a natural phenomenon consciousness is an integral part of the causal order, and by assuming consciousness to be ontologically distinct from the physical world substance and property dualism effectively rule out such causal integration[27]. But he thinks that consciousness is naturalisable despite its being ontologically distinct from the physical. His assertion is that conscious experience can be explained non-reductively, in contradistinction to standard scientific explanations which, as noted above, are reductive. However, as Chalmers points out, non-reductive explanations do occur in physics. These are explanations involving properties thought of as basic, or fundamental, such that they are not themselves reducible to simpler entities –

[27] Of course, property dualists suppose that phenomenal properties, i.e., qualia, *are* natural given that they are assumed to be in some sense caused by the physical processes in the brain even though they are causally inert. However, as I shall argue later, it is difficult to understand this epiphenomenalism to the extent that it is supposed that these phenomenal properties that arise from physical properties are causally inert *in toto*. Certainly a shadow, say, is largely epiphenomenal, but not entirely so – it has some

they are thus regarded as simple. Chalmers cites the example of James Clerk Maxwell's theory of electromagnetism (1995, 209). Before, scientists had tried, without success, to explain electricity and magnetism in terms of Newtonian mechanics. Instead, Maxwell introduced the concepts of electromagnetic charge and force, and postulated laws relating these properties to the other basic physical properties, these laws being expressed by four fundamental equations. This revised physics, with its expanded ontology, has successfully explained higher level phenomena. Chalmers summarises the general idea:

> Physical theories do not derive the existence of these features from anything more basic, but they still give substantial, detailed accounts of these features and how they interrelate, with the result that we have satisfying explanations of many specific phenomena involving mass, space, and time (1996, 213).

From this description of non-reductive theories we can discern two levels of explanation. First, at the more basic level, a "substantial, detailed account" of the basic properties themselves is provided by relating at large these properties, i.e., detailing how they interrelate. At the second level our understanding of these basic properties and the basic laws relating them is then used to provide explanations of particular higher-level phenomena.

Chalmers suggests that by thinking of consciousness as a basic property in the same sense as we think of the basic properties of physics we can, in similar fashion, construct a fundamental theory of consciousness. This proposed theory would "specify basic

---

causal efficacy however minimally. Qualia, on the other hand, are supposed to be entirely causally inert. This is very implausible. Moreover, because no other phenomenon is

principles telling us how [conscious] experience depends on physical features in the world" (1995, 210). Of course, by assuming consciousness to be a basic property, why it exists in the first place is left unexplained. Nonetheless, such a fundamental, i.e., non-reductive theory of consciousness at least promises to explain particular instances of conscious experience according to Chalmers. That is, it would enable us to explain what Chalmers calls "familiar phenomena involving experience" using principles concerning experience (*ibid*).

So according to Chalmers consciousness is naturalisable in that we can potentially explain particular conscious experiences in terms of this expanded ontology, i.e., in terms of all the basic properties postulated by physics *and* conscious properties also thought of as basic. A conscious property is a property of experience; qualia are taken to be paradigmatic conscious properties. Chalmers refers to such properties as 'phenomenal'. Thus we think of conscious properties in the same way as we think of gravitational or electromagnetic force, making no attempt to explain them in simpler terms, but rather to view them as irreducible in this sense. This is *prima facie* a promising idea since it appears to allow us to avoid any kind of supernaturalism vis-à-vis consciousness; that is to say, it entitles us to think of consciousness as a fully natural phenomenon, i.e., as nothing that requires postulating any supernatural causes. Let us look at this idea more closely.

Chalmers approvingly cites Saul Kripke's creation story (1996, 38;124;148). In his lectures *Naming and Necessity* Kripke likens consciousness to an afterthought of God. He imagines God creating the physical world, laying down, so to speak, all the laws relating the basic physical properties. In this world consciousness is absent. However, once this

---

strictly epiphenomenal in this sense we should wonder how qualia can be the exception.

purely physical world is established God *adds* consciousness to it, that is, God has more work to do in order to create consciousness in this physical world (Kripke 1972, 150-53). Consequently, the presence of consciousness makes no difference to the physical world in the sense that the laws of physics remain unchanged. Conscious properties, accordingly, can always be thought of independently of any physical properties. This construal of the relation between the physical and consciousness is endorsed by Chalmers. So although conscious properties are basic they are conceptually independent of the other basic physical properties. At most their relations to these basic physical properties are contingent, i.e., hold in the actual world but not in all possible worlds.

But, what is a basic property in this sense? It is essential to understanding the idea of a basic property that we define it in some way. Chalmers defines a basic property as one that is explanatorily simple, i.e., it cannot be reductively explained. But how do we determine if a property is explanatorily simple? If we do not know how to do this then we can only *presume* that some property is basic. Thus any property that *seems* simple in this sense is taken to be basic. This is not terribly helpful. In the case of basic physical properties we judge them to be basic because we can only understand them in terms of other basic properties. For example, in terms of traditional Newtonian physics at least, gravity is understood as the force related to mass, time, and distance expressed by the equation $F = Gm_1m_2/r^2$. In this sense a basic property is one that is essentially understood in terms of its relations to other basic properties. One cannot understand gravity or electromagnetic force without invoking other basic properties. This relates to Chalmers' remark, noted above, that physical theories "give substantial, detailed accounts of these [basic] features and how they interrelate" (1996, 213). One might argue that gravity, for

example, can be otherwise understood, e.g., as the force that makes things fall to the ground. But this is not to understand gravity as *basic*. By this measure it might turn out that this force can be explained in terms of simpler properties. Therefore, we know that a property is basic when it is understood strictly in terms of its relations to other basic properties, that together form a set of basic properties. Thus, gravity, electromagnetic force, mass, length, distance etc. together constitute a set of basic properties, where membership in this set is predicated on the property essentially being understood in terms of its relations with the other members of the set.

Conscious properties, however, are not members of this set of basic properties. They are not understood essentially in terms of their relations to these other basic physical properties. We do not think of conscious properties as basic in virtue of being an *integral* member of this set. But this is what defines these properties as basic. Its members are integral in that if one of the properties were removed from the set our understanding of the remaining members would be diminished. By this definition of a basic property, therefore, conscious properties are not basic. Certainly they are basic when defined purely in terms of being explanatorily simple. But again, this definition is too weak to be informative. So defined, conscious properties are basic insofar as they seem to be simple in this sense. Accordingly, we are only justified in supposing that consciousness itself is basic to the extent that it seems to be.

There are other reasons for rejecting the claim that consciousness is basic. In particular, as Chalmers admits, it would imply that experience is ubiquitous. That is to say, things can have conscious properties *unconditionally*, just as, for example, objects can have such basic properties as length or mass or gravitational pull for no underlying

reason – the object is not assumed to have these properties in virtue of realising even more fundamental properties. Thinking of conscious properties in this way clearly goes against our ordinary understanding of them. We assume, at least, such properties are chauvinistic biological properties, that is, we attribute them only to certain very complex systems such as higher animals. The idea that there is no reason to deny that a chair realises conscious properties strikes us as almost absurd. The construal of conscious properties as fundamental at least suggests that chairs can have experiences.

However, Chalmers bites the bullet and offers a spirited defence of the idea that experience is ubiquitous. This leads him to entertain some sort of panpsychism. In essence his response to obvious doubts about the possibility is twofold: First, he points out that we cannot rule out the possibility, i.e., it is not disconfirmable. No amount of observation of chairs, for example, can show that they do not realise conscious properties. Second, he tries to show that the idea is not absurd, i.e., it is plausible. He suggests that perhaps experience itself as such is not strictly ubiquitous, rather there might be even more basic conscious properties that in conjunction realise the kind of experience we have. So, for example, qualia are themselves constituted by even simpler or more basic phenomenal properties. In this respect chairs may lack the right combination of such properties that would enable them to have experiences as rich and complex as our own. He writes:

> ...perhaps there is some *other* class of novel fundamental properties from which phenomenal [conscious] properties are derived...Such properties would be related to experience in the same way that basic physical properties are related to non-basic

> properties such as temperature. We call these properties *protophenomenal* properties, as they are not themselves phenomenal but together they can yield the phenomenal...it is very hard to imagine what a protophenomenal property would be like, but we cannot rule out the possibility that they exist (1996, 126-27).

The first part of his defence does not of course demonstrate the truth of the idea. To attempt to do this would be to reason fallaciously, namely, to appeal to ignorance. The fact that there is no evidence to show that the idea is false does not demonstrate that it is true. Chalmers, no doubt, would accept this. Therefore it is the second part of his defence that must do the work, that is, show that the idea is plausible.

What is problematic about the hypothesis is not its plausibility, but the fact that it is *untestable*. This relates to Chalmers' claim that the hypothesis cannot be disconfirmed by observation (see *ibid*, 215). Above we noted that Chalmers remarks that it is very hard to imagine what a protophenomenal property would be like, but things are worse than this – these putative properties would be unobservable *in principle*. One might argue, in reply, that properties of subatomic particles like spin and charm are also unobservable, therefore protophenomenal properties are at least as respectable as these physical properties. However, these physical properties are observable in principle. Their existence is confirmable by observation, even if they cannot be directly detected. So long as thinking of subatomic particles as having these properties agrees with observation, i.e., fits with empirical evidence quite generally, we can suppose that they really exist. On the other hand, there is no method of confirming the existence of protophenomenal properties in the same manner. They are simply not thought of as empirical entities. Consequently, Chalmers' hypothesis can amount to no more than metaphysical conjecture.

That said, Chalmers tries to explain how conscious properties might be understood in terms of their relationship to the physical. Following Bertrand Russell he worries that basic physical properties are only understood relationally, especially in terms of their causal relations. This suggests the world is constituted by such relations. He remarks disapprovingly that "[t]he picture of the physical world that this yields is that of a giant causal flux, but the picture tells us nothing about what this causation *relates*" (*ibid*, 153). The proton, for example, is understood in physical terms simply as "that which causes interactions of a certain kind" (*ibid*). However, according to Chalmers, this understanding of the proton says nothing about 'what does the causing' as he puts it. I am not quite sure what more one could understand a proton to be exactly other than as that with such-and-such relations to other basic entities. It is not clear what is missing when we think of a proton in such relational terms. But Chalmers argues that such a purely relational understanding of basic entities leads to an insubstantial view of the physical world. It would not enable us to think of a proton, say, as having properties of its own. He holds that "it is more reasonable to suppose that the basic entities that all this causation relates have some internal nature of their own, some *intrinsic* properties, so that the world has some substance to it" (*ibid*).

So, very roughly, his thought is that our understanding of the world in terms of its basic physical entities and their relations leaves out a crucial aspect of reality – it presents us with a picture of the world as nothing but entities as nodes in a causal flux. Chalmers seems to think that the basic entities we name are substantive, that is, they are things in their own right and are not merely nodes, or 'placeholders' as he puts it. But any properties that basic physical entities have in this regard are by definition not relationally

determined. For that reason Chalmers chooses to call them intrinsic properties. Consequently, however, we have no idea about the nature of the intrinsic properties of the physical. Indeed, there would seem to be no way even to know about any such intrinsic properties. However, Chalmers suggests that phenomenal properties might plausibly count as such properties. Why *should* we think that this might be so? There is no reason, Chalmers tells us, except for the fact that the nature of these intrinsic properties "is up for grabs, and phenomenal properties seem as likely candidates as any other" (*ibid*, 154). There is to my mind at least something troublingly profligate about Chalmers' speculations above. He essentially posits an entire parallel set of properties which are independent of our physical understanding of the world. Again, this suggests that nothing we observe in the physical world could confirm or disconfirm the existence of these posited properties. The issue of metaphysical speculation was addressed in chapter 1. For now it is enough to note that while Chalmers' hypothesis that conscious properties are basic might be minimally plausible, the fact that there is no way of testing it makes it sterile.

Underlying Chalmers' suggestion that consciousness is best thought of as basic in some way is his claim that consciousness is ontologically irreducible to the physical. If he is right about this, then this suggestion makes sense at least to the extent that it allows us to find a place for consciousness in nature, i.e., to naturalise conscious experience in some way. He offers several arguments for this central claim, but the most important of them is his version of the conceivability argument. The form of this type of argument is that we can always think of consciousness as non-physical and therefore it follows that it

is non-physical. Next I shall consider his argument in detail, concluding that it is not persuasive.

2. 4 The Conceivability Argument

Earlier I remarked that consciousness seems incommensurable with natural, or physical, phenomena. At least, how we think of consciousness appears to be wholly distinct from how we think of physical phenomena. Some, such as Chalmers, take this incommensurability to show that consciousness cannot be a physical property. The basic argument is that if it is always possible to conceive of consciousness as distinct from the physical, then it cannot in fact be reduced to the physical. Therefore, consciousness is a non-physical property. A version of this argument was advanced by Descartes, for example, and as we shall see it has since been elaborated to strengthen its plausibility. Below is a formal presentation of the basic argument[28]:

> (1) Conceivability is an adequate test for possibility. If we can clearly conceive of its being the case that *p*, then it is possible for it to be the case that *p*
>
> (2) Where *x* is any conscious experience and *y* is any physical process, it is possible to conceive clearly of a situation when *x* is not identical with *y*.
>
> (3) If it is possible for *x* not to be identical with *y*, then it is false that *x* is identical with *y*.
>
> (4) Therefore, conscious experiences are not identical with physical processes.

Let us start with premiss (2). Few doubt that it is true. It seems that we can think of a person who, despite realising all the usual physiological states correlated with being in

---

[28] This is a paraphrase of Christopher S. Hill's presentation of the argument from conceivability, or what he calls the 'Cartesian argument' (see Hill 1991, 90).

pain say, nonetheless feels nothing. Nothing about realising such physiological states seems to force us to think that this person must feel pain. Turning to premiss (3): If, for example, it were possible to conceive of Cicero as distinct from Tully, then according to (3) Cicero cannot be identical with Tully. That said, even though it is in fact true that Cicero is Tully, it seems a possibility that Cicero is not the same person as Tully, suggesting that (3) is false. However, Cicero's being identical with Tully is necessarily true. This is because 'Cicero' and 'Tully' are coreferential terms. Consequently there is no possible world in which we can mean by 'Cicero' and 'Tully' different things. For, as Kripke points out, proper names refer rigidly. In other words, the meanings of proper names are fixed by their references. Moreover, their meanings must remain constant across all possible worlds – that is the only way we can meaningfully talk about the same object across possible worlds. It seems *prima facie* conceivable that Cicero is distinct from Tully, but only insofar as it is not known that the names are coreferential. Once the meanings of these names are known, it is impossible to hold that Cicero is distinct from Tully. Similarly, it seems possible, for example, to think of water as distinct from H_2O despite the fact that they are identical. But given that *in fact* these terms refer to the same substance, the only way to imagine them being distinct is to misdescribe water as being identical with some other substance, call it XYZ. So it appears that (3) is also true.

We are left with premiss (1). There is good reason, however, for thinking that (1) is false. We observed that it is not possible that Cicero is distinct from Tully because of the *fact* that we use these names to refer to the same person. But while this is how we actually use the names, there is nothing to stop us from imagining our using the names differently. Thus, there is a sense in which I could imagine that the Roman orator Tully

was never referred to as 'Cicero'. Therefore, I can conceive of Tully as distinct from Cicero. This is because, as Christopher Hill points out by way of another example, the concept Cicero is not logically identical with the concept Tully (Hill 1991, 92). That is to say, how we use the name 'Cicero' is not conceptually tied to how we use the name 'Tully' and vice versa. The same reasoning applies to the identity 'water is H_2O'. These concepts are likewise conceptually independent of each other. That is why we had to discover that they refer to the same substance, and as such the identity is an *a posteriori* truth. Given their conceptual independence we can imagine the world having turned out differently, such that the substance we call 'water' in virtue of its macrophysical, i.e., watery, properties was found not to be composed of H_2O, but of something else, call it XYZ. Clearly then not everything that we can conceive is possible, hence (1) must be denied. The conceivability argument fails, therefore, because the falsehood of (1) stops the inference from (2) to (3). That is, while it is conceivable that *x* is not identical with *y* it does not follow that it is possible that *x* is distinct from *y*.

That said, there is a sense in which what is conceivable is possible. We just noted that while in the actual world 'water' refers to H_2O, things could have turned out differently. This fact also expresses a possibility. However, if we think of possibility in this broader sense and the inference from (2) to (3) goes through, then we would have to conclude that because we can conceive of water as distinct from H_2O they are in fact distinct. This conclusion seems absurd. However, Chalmers offers a way of both holding (1) as true and rescuing the conclusion from absurdity, thereby defending the conceivability argument.

This defence centres on the distinction between two types of meaning, on what is called a "two-dimensional semantics" (see Chalmers 1996, 56-65). Chalmers follows Gottlob Frege who held that every concept has a sense [*Sinn*] which is said to determine the concept's reference [*Bedeutung*] (see Frege 1892/1997). Chalmers calls this sense the *intension* of a concept, which essentially is described by the function that relates the concept to its extensions. So, for example, the intension of gold is the function that relates the concept to its referent, namely, all the individual chunks of the substance. There are two ways in which intension can be thought to operate. We observed that in the actual world 'water' refers to the substance H_2O, and that things could have turned out differently. Another way of putting this is that in another possible world the concept could refer not to H_2O but to XYZ instead. The intension of the concept water in this sense describes a relation between the concept and its referent *relative to the possible world in question*. The idea is that the substance that has all the macrophysical, i.e., superficial, properties we associate with what we actually refer to as 'water' is not H_2O in some other possible world. This substance must be minimally thought of as 'watery stuff', namely, that substance that has the superficial properties we associate with actual water. If we imagine water distinct from watery stuff in this sense we have effectively stopped thinking about water *simpliciter*. This way of construing a concept's intension Chalmers calls its *primary* intension.

According to Chalmers premiss (1) of the conceivability argument is true when a concept is understood in terms of its primary intension. He asks us to consider the assertion 'water is watery stuff', stating that "we can know this statement to be true a priori, as long as we possess the concepts" (ibid, 64). By this he means that once we have

grasped the concept of water, i.e., understand what the term refers to, we immediatedly know that it is watery stuff, that is, no further observations are needed to know this. How is this so exactly? It is best to understand 'watery stuff' as an indexical term. In other words, however the actual world might have turned out, the term 'water' would refer to *that stuff around us* that has those manifest or superficial properties we actually attribute to water, i.e., its *watery* properties. Thus, in some other possible world 'water' may refer to XYZ and not H_2O, but insofar as we think of this substance as water it is because we assume that it is watery in this sense. This concerns the primary intension of the concept of water. Again, intension is thought of as that which maps the extension of the concept. And the *primary* intension maps the concept from any possible world to its extension, i.e., the class of things that the term refers to *in* that possible world. In terms of this primary intension we cannot conceive of water as not referring to watery stuff. And that is why when we grasp its primary intension we cannot fail to know that it is watery; that is to say, once in possession of the concept of water, and *a fortiori* its primary intension, we know *a priori* that it is watery stuff. This way of thinking of necessity must be distinguished from another concerning the concept's secondary intension. This intension maps the concept as it is determined in the the actual world to its extension *across* possible worlds. In the actual world water, understood as watery stuff, is in fact H_2O – and this is something which we have discovered about watery stuff, i.e., this is an *a posteriori* truth. Once what we mean by 'water' is fixed in this sense the term refers to H_2O in all possible worlds. This is to hold 'water is H_2O' as necessarily true.

The thrust of Chalmers' argument, as it relates to the conceivability argument more generally, is that in terms of a concept's primary intension conceivability is an

adequate measure of possibility, in accordance with premiss (1) in the conceivability argument above. We cannot conceive of water as not being watery stuff, i.e., thought of in terms of its primary intension, and *this* is evidence for its metaphysical impossibility. To use Chalmers' own terminology, watery stuff supervenes *logically* on water. Again, this is to say that if something is water then we also know *a priori* that it is watery stuff.

The important question is whether premiss (2) holds with respect to this revised conceivability argument. Still, is it possible to conceive clearly of a situation in which a conscious experience is not identical with some physical process? It is such conceivability that Chalmers considers to be the test of logical supervenience (1996, 93-122). If some pain quale supervenes logically on certain physical properties of the brain, i.e., neural properties, then when such neural properties are realised we should be able to know *a priori* that this pain quale is realised. But our understanding of conscious properties is such that we do not know *a priori* that a person realising such neural properties realises a pain quale. Indeed, as Chalmers argues, we can imagine a physical replica of a person, say of yourself, that lacks consciousness quite generally, i.e., your zombie twin. Knowing all the physical facts about a person does not entail knowing that she is conscious. Even with respect to particular types of conscious experience such knowledge is not entailed. Chalmers considers the intersubjective inverted spectrum scenario, where two people relevantly alike physiologically can still be imagined to have opposite colour experiences, e.g., I have a sensation of red looking at a ripe tomato while another person has a sensation of cyan, i.e., blue (1996, 99-101). It seems, therefore, (2) is true and conscious properties are not logically supervenient on the physical. By this

measure, the revised conceivability argument appears to succeed, and we should conclude that consciousness cannot be a physical property.

2.4.1 Reply to the Conceivability Argument

The objection to Chalmers' version of the conceivability argument that I shall look at, by Peter Carruthers, focuses on challenging his defence of premiss (1). Carruthers argues that because water and consciousness are natural phenomena it makes no sense to think of their concepts as mapping one-to-one onto their extensions as Chalmers assumes. Rather, the relation is many-to-one since a single natural phenomenon can always be conceived of in many ways. Therefore, our inability to conceive of a phenomenon in terms of particular properties does not entail its being metaphysically impossible that it reduces to these properties, which is what is assumed in premiss (1).

Carruthers argues that conceivability is not a measure of possibility, but for different reasons. Specifically, he argues that inasmuch as consciousness does not supervene logically on physical properties, contrary to Chalmers' claim, this has no metaphysical implications; that is to say, it does not show that consciousness is ontologically irreducible to physical properties. He notes that Chalmers countenances the belief that life is logically supervenient on the physical. Once all the physical facts about an organism are known, then we know *a priori* that it is living. There was a time, however, when it was commonly thought that life is an irreducible property, which some called 'élan vital'. But as Carruthers remarks: "They may have been mistaken, but they were surely not guilty of conceptual confusion, nor of mere failure to envisage the micro-physical realm in enough clarity and detail" (2000, 50). That is because the concept of

life was independent of the concepts concerning microphysical properties. And no amount of *a priori* analysis could have shown that life in fact reduces to such properties. Rather, people came to conceptualise life in terms of microphysical properties after a long process of scientific investigation. Thus, despite having originally thought of being alive as independent of any physical processes, this conception was not evidence of its being a non-physical property. Instead we take life to be real, i.e., a mind-independent property in the world – what Carruthers calls a 'worldly' property. Thus the term 'worldly' is used to connote a property's being 'out there' rather than being constitutively determined by how we think of it. How we think of a natural property makes no difference to the property as such in agreement with my precept. We understand that there are many modes of presentation of this property, and therefore many ways of conceiving of it. Properties are therefore individuated "thickly", as Carruthers puts it, that is, they can each be picked out by various conceptually independent descriptions. We have gradually come to understand how the descriptions of life in terms of its superficial properties and descriptions of the various processes, e.g., metabolism, in terms of microphysical properties are each descriptions of one and the same thing. It is by thinking of life as a worldly property in this sense that we are able to naturalise it, that is, to explain it in terms of other lower-level properties and the causal laws governing them.

Carruthers complains that it is because Chalmers does not think of consciousness as worldly that he concludes that it must be irreducible to physical properties. Moreover, this fact is indicative of Chalmers' construal of properties. According to Chalmers property terms are defined intensionally. Carruthers points out that Chalmers thinks of properties as mappings from worlds to extensions (2001, 54). Indeed, Chalmers states:

"We can see the intension of a property as a function of a world to a class of individuals (the individuals that instantiate the property), or from a world to properties themselves" (1996, 62). Consequently, he thinks that every discernible concept corresponds with a singular property; as indicated, for example, when elsewhere he writes that "such [primary] intensions will provide different functions from worlds to extensions (remember, there is just one space of worlds), and therefore distinct properties" (1999, 479). But, this is not to think of properties as real. It is instead to individuate them 'thinly', such that each property is said to be able to be picked out by descriptions involving a single concept alone. So, for example, suppose we conceive of two diseases *R* and *S* and we can coherently think of them as separate. By Chalmers' reckoning *R* and *S* are distinct properties. But if *R* and *S* are construed as worldly properties, the fact that they are *thought of* as distinct does not show them to be so. For as worldly properties it is seen as quite possible that some of their modes of presentation might be incompatible, such that they might in fact be the same property thought of in distinct ways. The only way to judge whether *R* and *S* are distinct or identical properties is again by empirical investigation, as in the case of the property of being alive. Yet Chalmers determines that consciousness is distinct from physical properties by appealing to the fact that we can coherently think of them as separate, i.e., that consciousness does not supervene logically on physical properties. That is fine, Carruthers notes, if consciousness is viewed according to Chalmers' construal of properties. But if we view consciousness as a worldly property then such thought experiments are of no use.

Carruthers acknowledges that many share Chalmers' intensional construal of properties. But he argues that to construe properties in this way is to abandon naturalism,

which holds that all properties are an integral part of the natural world governed by causal laws. More generally, he argues that thinly individuated properties are clearly distinguishable from real, i.e., worldly, properties. For example, he asserts that change strictly concerns real properties. He offers the following example: According to Chalmers the concepts of *grue* and *bleen* pick out distinct properties since they can coherently be thought of as distinct.[29] An object is said to be grue when it is green before the year 3000, say, and blue thereafter. Similarly, an object is said to be bleen if it is originally blue and turns green at the beginning of the year 3000. However, it is plausible to think of an object changing from grue to bleen at this precise time. There would, of course, be no discernible change in the colour of the object in question. From this Carruthers concludes that "*grue* and *bleen*, although perfectly legitimate as *concepts*, do not pick out real properties of objects" (2000, 36).

We are tempted to think that because consciousness does not supervene logically on physical properties, this is reason to suspect that it is a non-physical property. Carruthers argues in reply that this fact about consciousness does not show that it is non-physical, i.e., that it is not ontologically reducible to physical properties. Rather, this fact is simply a consequence of consciousness being a recognitional concept. He gives the example of chicken sexers to illustrate his point. There are apparently people who can sex chickens reliably without being able to explain how they do it. They sort the chicks into two groups, call them A and B, and these groups correlate strongly with male and female chicks respectively. This amounts to a recognitional capacity. It is relevantly analogous to how we grasp phenomenal concepts. We likewise grasp redness or pain, for example,

---

[29] This example is based on Nelson Goodman's grue paradox concerning the problem of induction (see Goodman 1965)

directly such that we do not understand them in functional or causal role terms, as we do with nearly every other type of concept. If the chicken sexer is asked can she imagine a world physically identical with ours but where A-type chicks are B-type chicks and vice versa, she is likely to answer that this is indeed conceivable, given that these types are not thought of in physical terms. However, Carruthers insists that this chicken sexer would be mistaken to conclude from this fact that these types are not reducible to physical properties. The same holds true with respect to consciousness. To the extent that consciousness does not supervene logically on physical properties it is because phenomenal concepts quite generally are recognitional. And this fact does not show that they do not reduce to physical properties.

Carruthers, then, rejects the claim that conceivability is a measure of possibility, namely, premiss (1) of the conceivability argument. He thinks that (1) is false because properties are real, i.e., worldly, and as such, how we conceive of them is independent of how they are in reality. That is, a property can be individuated by *distinct* concepts, hence what we may conceive of as distinct facts concerning a property may in reality be the same fact. In other words, what is conceivable may not be genuinely, or metaphysically, possible. It is worth noting that Carruthers presumes consciousness to be a natural, i.e., worldly, property. And in this respect his view differs importantly from that of Joseph Levine, another influential philosopher on this topic. Carruthers thinks that as a thickly individuated natural property, there is every reason to assume that consciousness is naturalisable in principle. It is no strike against the prospect of naturalising consciousness that it may not be directly reducible to the physical, that is, it may not supervene logically on physical properties. Indeed, he attempts to explain consciousness in terms of

intentional properties, which he holds, can in turn be explained naturalistically. Thus, given that reduction is transitive he thinks there is every possibility of at least indirectly explaining consciousness naturalistically.

Levine also argues against Chalmers' revised conceivability argument.[30] But Levine worries that Carruthers is simply dismissing the problem of the explanatory gap with respect to the possibility, that Carruthers envisages, of explaining consciousness naturalistically[31]. Carruthers essentially takes the explanatory gap to be the harmless . result of consciousness's being a recognitional concept. For example, he states: "While the 'explanatory gap' is of some *cognitive* significance, revealing something about the manner in which we conceptualize our experiences [i.e., in purely recogitional terms], it shows nothing about the nature of those experiences themselves" (1999, 67). It is this fact which explains why we cannot explain consciousness in physical terms. But since consciousness is nonetheless a natural property, it is still a candidate for reduction at least in principle. Levine also notes that, as it stands, consciousness can only be conceived

---

[30] Essentially Levine argues that we can understand concepts like water without concomitant knowledge of their manifest, i.e., superficial properties, *pace* Chalmers. Using the example of the concept of cat he states: "Of course it may be metaphysically necessary that cats are animals, but the crucial point is that it is not *a priori*. Mere competence with the term "cat" does not yield such knowledge." (2001a, 53) He does not deny that our understanding of concepts is often accompanied by knowledge of the salient properties concomitant with them, he only denies that it must include such knowledge. He calls this position "non-ascriptivism" given that it is to deny that we must ascribe *a priori* knowledge of a concept's concomitant properties with the grasp of its primary intension.

[31] The 'explanatory gap' is the term used to describe a major epistemological difference between identity statements concerning physical entities, e.g., 'water is H_2O', and psychophysical identity statements, e.g., 'pain is C-fibre stimulation'. Physical identity statements often help explain what something is, e.g., understanding that water is H_2O helps us understand why water has the properties it does. On the other hand, understanding pain as C-fibre stimulation in no way helps us to understand what pain is, i.e, to grasp the qualitative character of a pain experience. This particular concern is

recognitionally. However, he worries that there is no way of *showing* how consciousness is the natural, or real, property that Carruthers takes it to be. Levine, therefore, views the explanatory gap as a genuine epistemological difficulty for naturalistic theories of consciousness.

2.5 Summary

Each of the three arguments we have looked at are epistemological in tone, that is, at bottom their authors contend that consciousness does not ontologically reduce to the physical because of the seeming impossibility of *understanding* consciousness in physical terms. As we have seen there are many objections to this conclusion. And the success of these objections shows that the problem of consciousness is not sufficient evidence for the falsity of physicalism; at most our inability to understand consciousness in physical terms demonstrates that we can think of consciousness independently of the physical. But again, as we have seen, conceivability is not sufficient for metaphysical possibility, to paraphrase Levine. Insofar as it seems possible to think of pain separately from C-fibre stimulation it does not follow that they cannot *really* be identical.

But, I remarked in the introduction that I do not want to focus on the problem of consciousness as a problem for physicalism. Rather, my concern has been to evaluate the possibility of naturalising consciousness, that is, explaining it in terms of other natural properties. We have seen that the anti-physicalist arguments we have looked at are unsuccessful. They do not show that physicalism is false. From this fact we can infer that consciousness is at least in principle naturalisable. That is to say, there are no *a priori*

---

expressed by Levine in a forum discussion between himself and Carruthers (see Carruthers 2001 and Levine 2001b).

reasons for concluding that consciousness is not a fully natural phenomenon, i.e., a natural phenomenon among others, despite its peculiar epistemological status. And that is the conclusion I want to take away from this discussion. Importantly, however, I am not interested in defending a physicalist theory of consciousness to the extent that for any such theory it is assumed that qualia, as the properties of consciousness understood phenomenologically, are identifiable as particulars, i.e., they are individuable.[32] That is because I deny that they are individuable; and why this is so is the topic of the next chapter.

[32] This is not to imply that each of these theories are mutually exclusive.

Chapter 3

The Unindividuability of Qualia

In the last chapter we looked at the most influential arguments that repute to show either that physicalism is open to serious doubt or that it is outright false. In reply, I argued that none of these arguments succeed. Therefore, I concluded that there are no *a priori* reasons for supposing that consciousness is not physical, and consequently we can suppose that it is naturalisable, at least in principle. And it is the question of the naturalisability of consciousness to which I now want to return.

Earlier I acknowledged that our intuitive understanding of consciousness, i.e., our understanding determined by the first-person viewpoint, is ineliminable in that we cannot think of consciousness as such without understanding it in this way. In this respect we think of experiences in terms of their phenomenological qualities or qualia. For example, an experience of seeing something as green differs from seeing something as red because they have distinct phenomenological qualites, i.e., they realise different qualia. But it is one thing to postulate such properties of experience, it is another to make sense of them,

that is, to use the concept successfully to communicate facts about experiences. This way of thinking of experiences suggests that we can individuate qualia. The individuation of properties in general requires having a criterion by which we can judge when two instances of one of them are the same or distinct. In terms of qualia, there should be some way of deciding when one red quale, for example, is identical with or distinct from some other instance of a red quale. However, below I argue that no such criterion can exist. That is to say, qualia are unindividuable. As we shall see, the unindividuability of qualia implies that they are unnaturalisable, a fact that seems to contradict the conclusion above that consciousness is naturalisable. Later, in chapter 4, I shall argue that this contradiction is only apparent.

Below, in section 3.1, I begin by spelling out as precisely as possible the concept of qualia. There I suggest that, in agreement with Nagel's central assumption, qualia are perspectival in nature and that this fact is best understood if we think of qualia as constitutive of the phenomenal subject, i.e., the subject understood as that which has experiences. Then, in section 3.2, I argue that qualia thought of as constitutive of the subject are inapprehensible, that is, the subject cannot grasp them. Moreover, their inapprehensibility entails that they are unindividuable.

Next, in section 3.3, I consider Dennett's sceptical arguments against qualia. He contends that there are no such properties as qualia since the concept is incoherent. In arguing this Dennett likewise thinks that qualia are unindividuable, which he expresses in distinctive terms. I apply Dennett's insights in support of my claim that qualia are unindividuable. Here, like Dennett, I appeal to Wittgenstein's private language argument concerning why sensation terms cannot refer to features of experience that only the

subject is thought to have direct knowledge of. Crucially, I do not accept Dennett's conclusion that there are no such properties as qualia.

In section 3.4 I address some criticisms of Dennett's arguments that challenge his assumption that qualia are unindividuable. Initially I consider criticisms by Owen Flanagan, who argues that assuming our experiences are identical with certain brain states qualia, as properties of experiences, are straightforwardly individuable. Discussion of these criticisms is broadened to include arguments by David Papineau and Clyde Hardin, who similarly contend that qualia can be thought of as physical properties and consequently they are individuable. However, for similar reasons to Dennett I argue that the basic argument of all three of these philosophers is question-begging. Their reasoning is roughly as follows: they each present evidence for thinking that experiences are identical with brain states. And as properties of experiences qualia would thereby be physical. Moreover, if qualia are physical then they are individuable. Therefore, qualia are individuable. However, it cannot be assumed experiences are identical with brain states unless it is *already* assumed qualia, as the essential properties of experiences, are individuable to begin with, hence their argument is circular.

3.1 What Qualia Are Thought to Be

When someone reports that her tooth aches, for example, she aims for sympathy from her listener. Aches and pains, we all agree, are unpleasant feelings that we each want to avoid. Many of us have experienced a toothache and know first-hand how unpleasant it is. But, it does not seem to be necessary that her listener is able to identify with how she feels. Her intention in reporting a toothache is not to make him understand how her

toothache seems to her. In general, we assume there is a way sensations seem to others, as there is for ourselves. Further, we also presume that in some sense how a particular type of sensation seems to each of us is how it seems to others. But none of our everyday talk about sensations depends on this being the case; notwithstanding the fact that similarities in our behaviour in this respect suggest to us that toothaches probably feel the same way to each of us. But, again, our everyday sensation reports do not function as reports about what it is like to have such-and-such a sensation.

Many assume there is a fact of the matter about how a sensation seems to each of us. And if that is the case, then it makes sense to ask if a sensation seems to someone the same as or different from an earlier sensation, or even to ask if a particular type of sensation, e.g., a toothache, seems the same to two or more persons said to have it. That is, we can talk intelligibly about such seemings being alike or different from each other, both intrasubjectively and intersubjectively. The term of art for seemings is 'qualia'. Insofar as conscious experiences are thought of as a type of mental state, as distinct from propositional attitudes for example, we attribute to them certain identifying properties. These properties are their qualia.

Nonetheless, there is not complete agreement about the concept of qualia. Thus far we have defined them as the phenomenological qualities or features of conscious experiences, or what we have referred to as the "what it is like" of such experiences. When we perceive the sky as blue, for example, we do not simply recognise it as so, rather we feel it to be so. Paraphrasing Manuel García Carpintero, qualia might be minimally characterised as follows (2003, 357-58):

(1) Qualia, as properties, are essentially types.

(2) The realisation of qualia distinguishes the conscious mental sates of a subject from those mental states subjects have even when they are not conscious.

(3) Qualia are paradigmatically realised in mental states involving sensations, emotions, and imaginings.

(4) Qualia characterise what it is like for subjects realising them to be in those conscious mental states, the way things seem to them.

(5) Qualia are known by the subject realising them in a privileged way, in that subjects are said to know them non-inferentially, and consequently judgments about them are largely infallible.

Clearly (1) follows from thinking of qualia as properties of experiences, i.e., as types. (2) simply states that qualia are what make some mental states conscious, so to speak, as opposed to those mental states a subject might have while in a coma for example.[33] There is perhaps room for disagreement about (3). Some hold that propositional attitudes such as beliefs or fears have qualitative aspects to them, e.g., Flanagan. That is, there is something that it is like to believe that Napoleon was French, for example. Others, e.g., argue that propositional attitudes feel no particular way to their holder. With respect to (4), we have already discussed at length the characterisation of a quale as the "what it is like" of an experience when we considered Nagel's views in chapter 2.

Our focus will be on (5). Knowledge of qualia would seem to be privileged in that qualia are only apprehensible from a particular point of view. If I judge the pain I now

[33] Disagreement exists about this. In particular, while David Rosenthal (1986) agrees that qualia, or phenomenal properties, distinguish one sensory mental state from another, he argues that it does not follow that we are conscious of such a state and its phenomenal properties. In other words, he denies that qualia are properties that we are essentially aware of – we may not be conscious of their presence. What makes any such sensory mental state a conscious state, according to Rossenthal, is our *thinking about* it, i.e., attending to it. This view of qualia is not standard and the higher-order-thought (HOT) theory of consciouness on which this conception of consciouness is based is notoriously problematic (see e.g., Seager 1999, 60-84).

experience is like the one I experienced yesterday, it seems absurd to suppose someone else can dispute this claim. My judgment seems infallible in this sense. However, this infallibility might be spurious. We could imagine, say, a neurophysiologist identifying certain pain qualia with a specific type of neural activity. So, if you report having the same pain as you had yesterday but the neurophysiologist points out that according to his observations these pain experiences are qualitatively different, then he seems to show that you are *mistaken,* inasmuch as the identifications between pain qualia and types of neural activity are true. But it is not clear how this neurophysiological evidence can defeat your claim given that only you have access to the qualitative content of your pain experiences. In this sense your report is incorrigible, since others cannot apprehend your qualia directly. Another way of putting this is to say that if you cede authority to the neurophysiologist you might reasonably doubt your report, but if you do not doubt it then the neurophysiological evidence cannot *convince* you that you are mistaken. This issue is discussed in more detail later.

But, imagine the following experiment: A person is told that a hot object will be briefly placed on her skin – let us assume she is blindfolded – and when an ice cube is put there instead she reports a burning sensation. The experimenter immediately removes the blindfold showing her that the sensation she is having is not one of heat but of cold. This suggests that our first-person judgments about qualia are sometimes mistaken. However, this kind of misidentification of qualia is possible only to the extent that the sensations in question are alike. In this case the person is effectively reporting that she is having a burning-*like* sensation, and in this respect she is not mistaken. When a sensation of burning or cold are alike in their qualitative character what causes them is irrelevant. The

qualitative character of an experience, i.e., its quale, is not necessarily correlated with a particular type of property of an external object, just as, for example, we are supposed to be able to imagine two persons staring at a ripe tomato but realising distinct colour qualia according to the inverted spectrum hypothesis.[34] This fact about the way we think of qualia relates to the infallibility of our judgments about them. There is no way we can disabuse this person of her believing that she is having a burning-like sensation. In this sense her judgments are incorrigible. Likewise, if someone were to insist that how a ripe tomato seems to her is how grass seems to you, nothing we can do or say can convince her otherwise. To do that would require being able to compare these qualia publicly, and insofar as we must understand qualia as the ways things seem to each of us that is impossible.

3.1.1 The Peculiarity of Qualia

The characterisation of qualia in terms of judgments about them being infallible, stemming from our being acquainted with them in a privileged way, could perhaps be explained by their being intrinsic properties. While the extrinsic/intrinsic distinction is notoriously difficult to define, an intrinsic property is understood roughly to be a property an object possesses that does not depend on the object's relations to any other distinct object. But the very notion of intrinsicness, Dennett suspects, is spurious. He gives an example of a property we might think of as being intrinsic, namely, laughter. Dennett notes that we might explain laughter as an appropriate reaction to something funny, or more succinctly, to hilarity. However, this is not an explanation of laughter at all since hilarity in turn is defined as the cause of laughter. He likens such a pseudo-explanation to

[34] For a detailed discussion of the inverted spectrum hypothesis see Shoemaker 1996.

that given by Argan in Molière's play *La Malade Imaginaire*, who, when asked why opium puts people to sleep, responds: 'because of its *virtus dormitiva'*, i.e., its dormative power (see Dennett 1991, 63). The best we can do to define laughter is by such circular means because it is an intrinsic property. As evidence of the spuriousness of the concept of intrinsicness he points out that not even a brilliant theorist like David Lewis has succeeded in explaining the intrinsic/extrinsic distinction (see Dennett 1986, 67). Calling some property *intrinsic*, Dennett contends, follows from the desire to think of it as being essentially so. In the case of laughter, for example, we think of it as 'just so' in that if we were to explain it in terms of relations to other objects and their properties, we are strongly inclined to believe that what we essentially mean by the term 'laughter' would necessarily be omitted. But in general he thinks claiming that certain properties are intrinsic on these grounds is simply question-begging (*ibid*, 68).

But, the idea that qualia are intrinsic points to something very important, namely, that qualia are thought of as being *essentially* perspectival or subjective. What is peculiar about the concept, i.e., how we think of qualia, is that they are inapprehensible to others. As Nagel noted, we think of them as being dependent on a particular perspective. However, we think of properties as universals or as types, and as Carpintero's list indicates, we think of qualia in this way as well. But, if a red quale for example is a type of phenomenological quality that can be tokened by other people, then what does it mean to assume that qualia are subjective, i.e., *dependent* on a particular viewpoint? The only way to make sense of qualia as types, it would seem, is to think of them as *independent* of the viewpoint of the subject they concern. After all, we do not want to assert that a quale is the phenomenological quality of experience as realised by a particular subject *only*. If

we were to hold this then we could no longer coherently think of qualia as properties. In order to think of qualia as properties, i.e., as types, while maintaining that they are perspectival by nature à la Nagel, I suggest that we think of them as *epistemically* private. That is, they are in some crucial sense constitutive of us as phenomenal subjects. Indeed, it is the attribution of these properties to experiences which enables us to think of experiences as distinctive phenomena in some sense. What distinguishes pain from mere neurological activity and certain behaviour? It is its phenomenological quality, i.e., its subjective aspect. We might express this thought by the following principle:

(S) No two subjects can *know* that they are realising the same type of quale.

By this measure, it would be impossible for distinct subjects to know that they are realising the same quale-type. Here we think of the subject, minimally, as that which has experiences.

The obvious worry about this claim is that, as noted in our discussion of Jackson, it leads to a radical scepticism about other minds (see section 2.2). If I can never know that you or anyone else is having the same type of qualitative experience as me when, say, experiencing a toothache, then I cannot rule out the possibility that everyone else is in fact a phenomenal zombie. However, (S) does not necessarily entail that there is nothing that it is like to be another subject, rather it states that no subject can ever know if someone else is having the same *type* of qualitative experience. Accordingly, (S) is consistent with assuming that there is something that it is like to be another subject.

Moreover, as I argue later, there are no grounds for assuming that other normal human beings lack phenomenal consciousness.

More generally, I think it is helpful to distinguish between direct *acquaintance* with qualia and *knowledge of* them. Assuming that acquaintance is a species of knowledge very broadly construed, it is nonetheless important that we understand the differences between these kinds of knowing. Earlier, in my discussion of Jackson's views, I said that something is knowable if and only if it is knowable by others. Here we are talking about propositional knowledge, as opposed to mere acquaintance. Indeed, I argued that it is the equivocation of these distinct ways of knowing that led Jackson mistakenly to conclude that qualia fall outside of our physical knowledge about the world. And we have this kind of propositional knowledge so long as it is justifiable, and as such it is inferential in nature. But *acquaintance* with our own qualia is non-inferential. Indeed, we cannot justify to others how we know, for example, that we are realising a red quale. Moreover, given that knowledge is usually defined as justified true belief we might conclude that talk of knowing that we realise such-and-such qualia is not possible, i.e., it makes no sense. I shall say more about this later.

3.2 Why Qualia Cannot Be Individuated

The epistemic privacy of qualia implies that they are not apprehensible by others. In fact, I contend, they are not apprehensible at all since we cannot apprehend our own qualia either. Qualia are constitutive of us as phenomenal subjects; that is to say, it is in virtue of realising qualia that we are subjects. One might say that the phenomenal subject realises itself as qualia. Another way of understanding the perspectival nature of qualia in

this sense is to think of them as properties by which we apprehend things in the world. This fact explains the transparency of qualia and conscious experiences more generally. We cannot apprehend qualia by some sort of introspection of our perceptual experiences, rather we *see right through* them so to speak to the external properties of the things the experience is of. Michael Tye offers this helpful appraisal of our situation:

> Try to focus your attention on some intrinsic feature of the experience [looking at a blue painted square] that distinguishes it from other experiences, something other than what it is an experience *of*. The task seems impossible: one's awareness seems always to slip through the experience to the blueness and squareness, as instantiated together in the external object (1995, 30).

The inapprehensibility of qualia is thus manifested in their transparency.

Crucially, the inapprehensibility of qualia implies that there is no distinction to be made between an experience having a particular quale and an experience *seeming* to have this quale, since seeming to have a quale would entail apprehending it.[35] This is impossible because a quale, as an integral part of the subject, cannot itself be

---

[35] This basic point is made by Saul Kripke in his *Naming and Neccesity*, where he states that "in the case of mental phenomena there is no 'appearance' beyond the mental phenomenon itself" (1980, 154). It is this characteristic of mental phenomena, e.g., qualia, that Kripke thinks precludes experiences being identified as physical states. According to Kripke this is because there could be no way of reconceiving qualia so that the ostensible distinctness of the experience identified by some quale from a certain type of brain state is ruled out. If we can conceive of an experience as distinct from a certain type of brain state we cannot be mistaken about this since there is nothing about its quale beyond this appearance. The ways things appear to us cannot themselves be only apparent, i.e., possibly otherwise.

apprehended by the subject it is part of, just as the eye cannot gaze upon itself. This fact precludes the possibility of individuating qualia.

To individuate a property is to determine if some instance of it is distinct from or identical with some other instance. Individuation requires possession of a criterion, a standard by which to make such a determination or judgment. It involves using a criterion to determine whether a thing that seems to have some property does have that property rather than some other. We need such a criterion to decide, for example, whether a sofa is maroon in colour. Unless a criterion exists we would effectively be unable to predicate *maroon* of anything. So, for example, in the case of the property of being a certain height, e.g., one metre, we determine this by devising a standard scale to make a direct comparison between the apparent height of an object and some mark on this scale. Thus, having a scale enables us to judge if the object is one metre high, i.e., has this property. Here we distinguish between seeming to be one metre high and being one metre high. If there were no such distinction so that we took the height of an object to be whatever it seems to be, then different instances of being one metre high would be incomparable. We could not determine if two objects have the same property or distinct properties, i.e., distinct heights.

But this is the situation with respect to qualia. There is no gap between a quale and how a quale seems to be. So, we cannot determine if the quale of my toothache is distinct from or identical with the quale of your toothache, for example, because there is no possibility of comparing them. In general, therefore, qualia are unindividuable, that is, there cannot be any criterion by which we can judge when one quale is either distinct from or identical with some other quale. An important upshot of the fact that qualia are

unindividuable is that qualia terms cannot refer, since by any such term we do not identify a particular property.

But while perhaps we cannot compare qualia intersubjectively in the way suggested by the example of different people's toothaches, still we might think that qualia can be individuated from a first-person viewpoint. The idea would be that intrasubjectively we are able to *judge* whether one instance of a quale is the same or different from another across time. We each feel confident, after all, that we know that how the colour of a lawn seems to us one day is the same as it seems to us the next day, for example. There is, however, a fatal difficulty with this idea. It is impossible that such an intrasubjective exercise could amount to a judgment; there can be no private judgments in this sense. To make a judgment requires the possibility of distinguishing how something seems to someone and how it actually is, and this distinction is impossible from a single point of view – there is no sense in which the subject making the so-called judgment could make such a distinction.

This point is famously made by Wittgenstein vis-a-vis rule-following. To follow a rule is to obey it. But, Wittgenstein observes that "to *think* one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule 'privately': otherwise thinking that one was obeying a rule would be the same thing as obeying it."[36] This is because obeying a rule is a practice, i.e., something we do that is necessarily 'done in the open'. To make a judgment is a practice in the same sense. My judgment that it is raining, for example, is only so to the extent that there is a distinction between its seeming to me that it is raining and there being a fact or not as to whether it is raining. Accordingly, for someone to judge that she is realising a red quale, say, could be a judgment only if she can make the

distinction between how it seems to her that she is realising a red quale and whether or not she is actually realising a red quale. However, qualia are defined as how things *seem* to the subject. Therefore, the utterance 'I am realising a red quale' is empty since there is no distinction between how red seems to someone and how red 'really' seems to her. Accordingly, for someone to claim that she is realising the same colour quale as she did yesterday cannot amount to a judgment.

3.3 Dennett's Qualia Scepticism

In supposing that qualia are unindividuable I take it that it is this fact about them that marks them out from ordinary properties. But how can we think of qualia as properties at all if we cannot in principle distinguish one instance of such a property from another? This question brings us to Dennett's views. While Dennett does not use the term 'individuability' he essentially argues that qualia are unindividuable from the third-person viewpoint, and that consequently the concept is incoherent. He likens the concept of qualia to such bogus scientific concepts as élan vital and caloric. These properties were originally posited by theories used to explain certain phenomena, namely, life and heat respectively, but were eventually dropped when the theories they concerned were replaced by predictively more successful ones that did not posit them. The theories employing the concepts of élan vital and caloric came to be seen as false and so the concepts were judged to have no extensions; we came to realise that there were no such properties. Dennett argues that psychological explanations of our behaviour that invoke qualia are empty in the sense that they are empirically untestable. Our judgments

---

[36] Wittgenstein 1958, #202.

concerning qualia are unconfirmable because of how we think of qualia, i.e., as intrinsic properties that are unobservable from the third-person viewpoint. This fact suggests that like caloric the concept of qualia is useless. For example, positing them does not help us to explain the behaviour in the way that it is sometimes appealed to. Overall, there is no reason to think qualia exist according to Dennett.

I disagree with Dennett's conclusion that qualia do not exist. My reasons for disagreeing will be given later, but for now it should be noted that there is the straightforward objection to it that qualia self-evidently exist. Given that we have phenomenal experiences it seems absurd to suppose that they do not have this peculiarly perspectival aspect to them. Denying the existence of qualia seems to amount to denying this fact about experience. This concerns what I previously referred to as the ineliminability of our phenomenonological understanding of consciousness. But what is useful about Dennett's arguments, notwithstanding the seeming absurdity of his conclusion, is his showing how it is that qualia are unindividuable. And it is this aspect of his arguments that I want to focus on.

His arguments start off from Wittgenstein's observations about the nature of sensation talk. Following Dennett it is worth considering some aspects of Wittgenstein's renowned private language argument since the argument is directly relevant to the impossibility of individuating qualia. In his article 'Quining Qualia' Dennett compares his position to that of Wittgenstein (Dennett 1988). He suggests that Wittgenstein effectively denies that there are such things as qualia, citing the following well-known passage from his *Philosophical Investigations* in which Wittgenstein argues that a sensation term cannot refer to some private, i.e., inner feature, namely, what we are here calling the qualitative

character of an experience. He compares such inner features to the privately accessible content of a box. He writes: "The thing in the box has no place in the language-game at all; not even as a *something*; for the box might even be empty. – No, one can divide through by the thing in the box; it cancels out, whatever it is" (Wittgenstein 1958, 293). It is important to be clear what Wittgenstein's claim is here.

According to his early philosophical views, as expounded in *Tractatus*, the essential purpose of language is to describe the world, i.e., the facts. As such language, he thought, mirrors the logical structure of the world. Later, however, he came to view language in a different way. He began to see it in terms of an activity, as a practice; or more precisely, as a myriad of practices, reflecting the enormous variety of uses to which we put language. Thus, language is a complex of particular practices each of which can be described as a "language-game". Here the term "game" is used to capture the sense in which the practice is characterised by following rules. By this measure our concepts, as determined by such rule following activities, are constituted by the language-games in which they are employed. A concept is best understood in terms of how it is used in specific language-games.

In the passage cited above Wittgenstein focuses on the language-games concerned with how we report to others about states internal to us, that is, states that others cannot directly observe, e.g., a decaying tooth or a sprained wrist. These are states that, of course, we each come to know about through particular sensations. Often in such reports we relate to others what kind of sensation it is that we are having, e.g., "I have a headache." Thus, we might be tempted to suppose that by such reports we are not only expressing to others our discomfort, pleasure etc., but also we are telling them something

about the sensation *as identified by its qualitative character*, i.e., its quale. Wittgenstein, however, denies that our reports can ever be about such qualia. To explain his position he likens the qualitative character of sensations like toothaches to the content of a box, content that *only* the owner of the box can see or apprehend. He imagines a whole community of such box owners, all of whom call the content of their box a "beetle." He observes that since no one can apprehend the content of anyone else's box, the term "beetle" cannot be used to refer to it, i.e., the thing inside each box. The term "beetle" can only be used by the members of this community to talk about the content *simpliciter*, i.e., that which is inside the box, *whatever* that happens to be. Consequently, what happens to be inside any one of the boxes does not affect how the term "beetle" is used; indeed, as Wittgenstein remarks, the box could even be empty. Hence his insistence that the thing inside the box can play no role in this language-game. Our language-games concerning reports about our internal states are analogous to that of the imagined box owners, our sensation words cannot be used to refer to the qualitative character of the sensations. Qualia are relevantly analogous given that no one else can apprehend the one we each realise.

However, Wittgenstein is not claiming that sensation words like 'pain' are about nothing, and to stress this point he adds that while the thing in the box "is not a *something*, it is not a *nothing* either!" (*ibid*, 304). Thus, strictly speaking Wittgenstein does not deny the existence of qualia outright, as Dennett initially suggests. At most he is suggesting that our sensation words do not refer to any such inner features.

But regardless of the conclusion Dennett aims to defend, his arguments are based on denying the possibility of individuating qualia from the third-person viewpoint. And with

this aim in mind he presents what he calls intuition pumps, i.e., thought experiments "designed to flush out – and then flush away – the offending intuitions" (1988, 44). These offending intuitions are those that lead us to think of qualia as ordinary properties in the sense that we can treat them as identifiable particulars. Dennett insists that this is not possible. One of the characteristics he identifies qualia as having is immediate accessibility to each of us. That is, we are only acquainted with qualia from our own case. For example, I am acquainted with what it is like to be in pain, i.e, to realise a pain quale, through *my* experiencing pain. I could not have gained this knowledge through anyone else's experience of pain. In other words, we know qualia non-inferentially such that our knowledge of qualia is secure.

Dennett sets about showing how thinking of qualia in this way is incoherent. By describing qualia as an incoherent concept I interpret him as claiming that, unlike in the case of concepts concerning ordinary properties, judgments involving qualia are unconfirmable, making it impossible to employ the concept in the same manner as we standardly employ other concepts. He considers the example of intersubjective spectrum inversion, asking you to imagine waking up one morning to discover that colours seem inverted to you, so now the sky looks yellow, the grass has a distinctively magenta hue, and so on. You have reason to suspect that some unethical neurosurgeon has tampered with your brain to effect the inversion of your colour qualia. Dennett notes that the neurosurgeon could have done one of two things to you (*ibid*, 50-51):

(1) Your optic nerves have been altered so that "all relevant neural events 'downstream' are the opposite of their original and normal values."

(2) Your memory-access links have been inverted, causing you to remember everything as having the opposite colour to that which it presently seems to have.

Both of these procedures will lead you to think that your colour qualia have been inverted. However, only (1) is supposed to cause such inversion. In the case of (2) your colour qualia are assumed not to have changed – it is your *memory* of how colours seem to you that has been altered. Dennett observes that there is no way for the victim of this surgery to know whether her colour qualia have been inverted *for herself.* In order for her to know this she would have to seek outside evidence. Certainly a neurophysiologist might be able to confirm whether she has undergone procedure (1) or (2), but from the victim's perspective there is no way of determining which procedure she has undergone despite its being supposed that only (1) leads to the inversion of her colour qualia.

However, it might be objected that the neurophysiologist only confirms whether her *memory* has changed. As Seager puts it: "The question should be: after the apparent inversion, do I have any reason to doubt my memory?" (1999, 99). Thus, after waking up you are equally concerned to know whether your memory has changed or not. And it is this concern that the neurophysiologist can answer. But it is not clear that this is a successful reply to Dennett's attack on qualia. Let us suppose that the neurophysiologist confirms that it is your memory that has been tampered with. What conclusion can you draw from this? You might conclude that the cloudless midday sky, for example, seems to you the opposite colour that you remember it seeming, but in reality your colour qualia are not inverted, so the sky still seems blue to you – you just falsely remember it seeming to be a different colour. However, this is plainly contradictory. The sky does not seem to you to be the same colour, for if it did you would have noticed nothing after waking up.

This objection presupposes that how colours seem to us is independent of how we remember them seeming. The sky still seems blue to you, it is just that you misremember it as once seeming yellow. But, what colour would the sky seem to you after the procedure described in (2), blue or yellow? The difficulty is that it is impossible to imagine this. We cannot think of how colours seem to us as independent of how we remember them seeming; and this relates back to the fact that qualia are unindividuable.

3.3.1 Qualia Debunked

Another of Dennett's thought experiments concerns a story about two professional coffee tasters, Chase and Sanborn. Each gives a different explanation for why, after many years with the company, he no longer enjoys the taste of Maxwell House coffee:

> Chase: While the coffee has the same flavour I have become a more sophisticated coffee drinker and no longer like it.
>
> Sanborn: My tastes have not changed, but my taste buds have and as a result I no longer like the coffee either.

Note that neither man claims that the coffee's flavour has changed. Chase, then, contends that his taste qualia with respect to the coffee have remained the same, whereas the effect these qualia have on him has changed. Sanborn, on the other hand, thinks that it is his taste qualia that have changed, and if they had not he would still enjoy the coffee's flavour. The important thing to note is that both men's accounts of why they no longer enjoy the company's coffee appeal to qualia, and as such they assume qualia to be entities in their own right; that is, their existence is what makes the difference. But the problem with these accounts, Dennett suggests, is that they are ultimately unconfirmable.

Accordingly, appeal to qualia serves no purpose; thereby indicating that the notion of qualia is empty.

Why should Dennett think that these accounts are unconfirmable? Because, he argues, there is no fact of the matter as to which is true. But how does this follow? It seems plausible to suppose that there is a fact of the matter even if we cannot confirm which is true. Indeed, Dennett himself notes,:"[i]t seems easy enough... to dream up empirical tests that would tend to confirm Chase and Sanborn's different tales" (1988, 55). For example, Chase may fail to re-identify coffee and other drinks when blind-folded, with only minutes between sips. His failure would of course strongly undermine his claim to know that the taste of the comapany's coffee has remained constant over the years. However, Dennett chose the example of *two* tasters to characterise two polar opposite accounts of the same phenomenon, namely the gradual loss of enjoyment of the coffee's flavour. In such cases these kinds of empirical tests might indeed seem persuasively to confirm Chase's or Sanborn's stories. But here we overlook another explanation of this change. It is equally plausible to suppose that their loss of enjoyment might be explained in terms of a combination of the two factors each cites as a singular cause. That is to say, both Chase's and Sanborn's loss of enjoyment could be explained as being the result of an imperceptible shift in their taste qualia *and* a shift in their reactions to these qualia. This account, however, cannot be confirmed conclusively by any such empirical tests. In other words, empirical tests could only confirm the kind of extreme accounts represented by the examples of Chase and Sanborn.

Nonetheless, there seems to be another way of confirming such accounts empirically, and that is to appeal to changes in neurophysiology. Dennett imagines, for instance,

Sanborn appealing to a change in his taste organs to explain the change in his taste qualia. Similarly changes in Chase's neurophysiology might confirm his account. This kind of empirical confirmation at least seems possible in principle. But, Dennett contends that there is a limit to the evidential power of neurophysiology as well. To make this point clear, in another thought experiment, he presents the following scenario: We are asked to imagine that Chase undergoes a surgical procedure that results in the reversal of his taste qualia; so, for example, things that once tasted sour to him now taste sweet. But over time Chase compensates for this reversal so that eventually there is no discernible difference in his behaviour, including his linguistic behaviour, i.e., he learns once more to call ice-cream sweet etc. His readjustment is so complete that he cannot remember how things tasted any differently before the operation from how they do now. If the neurophysiological changes correlated with this readjustment occur early in the process, then we might plausibly suppose that Chase's qualia have been restored to how they were before the operation. If, on the other hand, the correlated neurophysiological changes occur later on in the process, then we might suppose that Chase's qualia remain as they were immediately after the operation but he mistakenly remembers them as being the same as they were before the operation. Here we suppose that the changes have occurred in the part of the process concerned with memory, where in effect the qualia, as putative properties, act as inputs in the memory process. The trouble is, Dennett points out, we have no way of determining which of these two possibilities is actually the case. Chase cannot know what is really the case by appealing to how things seem to him, since as Dennett observes:

> Chase may think that he thinks his experiences are the same as before [the operation] *because* they really are (and he remembers accurately how they used to be), but he must admit that he has no introspective resources for distinguishing that possibility from [the] alternative... on which he thinks things are as they used to be *because* his memory of how they used to be has been distorted by his new compensatory habits (*ibid*, 58).

Now, we might appeal to neurophysiological evidence to decide which is the case. However, Dennett asserts that no neurophysiological facts can ever tell us where in the process the qualia effectively are formed such that they become a factor in this sense. In other words, there can never be any neurophysiological evidence that allows us to *decide* the case. The concept of qualia he is attacking, recall, is of properties that are apprehended immediately and *only* by the subject. Imagine in Wittgenstein's beetle-in-a-box scenario that whatever is in each box is always correlated with certain phenomena, for example, every box rings simultaneously. This seems like good evidence for supposing that each box contains the same type of object, even if we cannot directly confirm this assumption given that we cannot view the things in other people's boxes. But it is crucial to note that the content of each box is essentially private; that is to say, there is *in principle* no possibility of anyone other than the owner being able to view its content. Therefore, it simply makes no sense to talk of the things in the boxes as being the same thing, ditto for qualia. As Dennett remarks: "The idea that one should consult an outside expert, and perform elaborate behavioural tests on oneself to confirm what qualia one had, surely takes us too far away from our original idea of qualia as properties with which we have a particularly intimate acquaintance" (*ibid*, 60). While this follows if

qualia are thought of as essentially private, as we shall see, some contend that qualia do not have to be thought of in this way.

3.4 A Physicalist Criterion of Individuation

A common response to Dennett's position is to argue that qualia are identical with certain brain states, a view which of course goes back at least to U.T. Place's and J.J.C. Smart's defences of the mind-brain identity thesis. Accordingly, the concept of qualia is straightforwardly coherent and their existence is not in doubt. A detailed version of this line of argument is presented by Owen Flanagan in his book *Consciousness Reconsidered* (1992). Flanagan endorses Dennett's description of qualia as the ways things seem to us. However, he adds that this "characterization is useful so long as we do not join it to some contentious theory of qualia according to which first-person seemings exhaust the nature of states with a seeming aspect" (1992, 62). And this is precisely what he accuses Dennett of doing. He argues that Dennett's construal of qualia is too narrow in this sense. Flanagan's claim is curious given that qualia are defined, at least, as those properties of experiences apprehended from the first-person point of view. How, therefore, does he suppose them to be *more* than first-person seemings? For one thing he thinks of qualia as more in terms of mental states than as properties. Rather, he asserts that a quale "is a mental event or state that has, among its properties, the property that there is something it is like to be in it" (*ibid*, 64). However, the description of this property of a quale, that it has among others, itself fits exactly with the general description of qualia. In this respect, what Flanagan takes qualia to be is difficult to discern. But, in general, he holds that qualia in a wider sense, as he puts it, represent a typology of mental states, that is, the

term describes the kind of mental states that have a qualitative aspect to them, i.e., they seem some way or other to us. These can include propositional attitudes according to Flanagan.

Flanagan is sympathetic toward some sort of identity theory; he thinks that it is at least a plausible empirical hypothesis (see *ibid*, 56-60). Thus, he suggests that "[s]ubjective experiences have particular spatial locations in the form of distributed neural activity. The neural properties of qualitative experiences are not revealed in the first-person point of view" (*ibid*, 64). In agreement with Paul Churchland, he is inclined to type-identify colour qualia with "a specific triplet of spiking frequencies in some triune brain system" (P.M. Churchland 1989, 104). Flanagan, therefore, holds that qualia can be more broadly conceived as mental states with a qualitative aspect, but which also have physical features to them.

To make clear the contrast between his view and that of Dennett Flanagan cites an analogy employed by Dennett. Dennett compares our supposed conceptual muddle with regards to qualia to the snarled string of a kite, where in the case of such a kite there comes a point when it is simply easier just to get some new string. He writes: "That's how it is, in my opinion, with the philosophical topic of qualia, a tormented snarl of increasingly convoluted and bizarre thought experiments," (Dennett 1991, 369). Flanagan, on the other hand, contends that the problem is better likened to the strings of two kites getting entangled. The one kite represents an uncontentious conception of qualia – what he calls the box kite – and the other represents the problematic conception based solely on the subjective aspects of qualia. He declares that the solution is to disentangle these kites and to discard the problematic one, leaving the box kite to "fly

freely" (1992, 63). This uncontentious conception of qualia is to think of them simply as the "ways things seem to us" which, as noted earlier, is a characterisation offered by Dennett himself.

Addressing Dennett's thought experiments discussed in his 1988 article "Quining Qualia," Flanagan specifically considers the example of Chase's surgical inversion of his taste qualia. Dennett lists two hypotheses that could explain how, after six months, Chase reports that his taste qualia have returned to how they were before surgery, namely, (i) that his taste qualia are still abnormal, but changes in memory-access processes in his brain during the six months has altered his memory, and (ii) that somehow the changes in his brain are such that his memories are compared to relevant neural activity prior to, or upstream from, the qualia phase of his brain processes, so he remembers his old taste qualia again. As noted, Dennett claims that we cannot decide between the two hypotheses, no matter how much empirical evidence there might be in favour of one over the other. That is because, according to Dennett, the reasons a neurophysiologist might have to prefer one hypothesis over the other follow only from her appropriating the term "qualia" to her own theoretical ends. Flanagan confesses that he is confused by Dennett's claim (1991, 77). He interprets Dennett as assuming that the reason why no empirical evidence can decide in favour of one hypothesis over the other is that he simply denies that the term "qualia" refers to anything at all, given his narrow conception of qualia. In reply to Dennett's story Flanagan imagines that upon watching a video of himself just after his operation Chase remarks to his doctor that he has now come to realise that, six months later, his taste qualia are back to normal. The doctor agrees with his judgment and offers him two plausible hypotheses for explaining in neurological terms how this

readjustment has come about. The doctor explains how his team settled on one rather than the other. Flanagan remarks:

> The case at hand is one in which Chase knows that something is happening – he can describe it up to a point – but he doesn't understand what is happening in any depth, so he rightly cedes authority to the experts. It violates the logic of the concept of qualia as Dennett construes it for first-person authority to be ceded in this way. To think this, Dennett must think that the identification of qualia with the "ways things seem to us" must be interpreted as meaning that a quale is a state such that, necessarily, being in it seems a certain way and, necesssarily, there is nothing else to it. But...I claim that the concept simply doesn't need to be understood in that way (1992, 79).

However, a crucial part of Dennett's complaint about the notion of qualia is that how we think of them, i.e., as intrinsic properties, makes them conceptually independent of physical properties. In essence Flanagan denies that qualia are intrinsic, that is, he supposes that we can understand them to some degree in terms of their relations to other properties. Nonetheless, he admits that he does not know how we can understand qualia, as subjective phenomena, in physical terms. With respect to the question of how subjectivity is part of the material world, he admits to being unable to offer any answer. Rather, he argues that given the immense complexity of the brain it is not implausible to expect that subjectivity can arise from it (*ibid*, 60).

If Flanagan is right and qualia are identical with brain states, then we cannot assume that qualia are unindividuable after all. So we are presented with a theory of

consciousness that seems to undermine principle (S), namely, the assumption that only the subject can be acquainted with the qualia it realises. Flanagan's overall position is that while we do not know how qualia, as subjective phenomena, are natural, there is no reason to suppose that they are not so. Intuitively we think of qualia and brain states as distinct since the former, but not the latter, is subjective in nature. David Papineau acknowledges this widespread intuition. Nonetheless, he argues that we have good reasons to reject it. However much we might *feel* that they are distinct, to believe that they are is too problematic. Moreover, he recognises that how we think of experiences from the first- and third-person points of view is distinct. But, the fact that we think of them in distinctive ways does not imply that we are thinking about distinct phenomena. To reach this conclusion is to commit a fallacy, what Papineau calls the "antipathetic fallacy"[37] (1993, 114-18). So, even though we cannot explain how the subjective aspect of qualia relates to physical properties, it does not follow that qualia are not identical with brain states. Of course, here Papineau is essentially making the same point as those who oppose the conceivability argument, such as Levine and Carruthers. That is, he is pointing out that how we conceive of properties must be distinguished from the properties themselves, i.e., from what Carruthers calls "worldly" properties. What is more, Papineau contends that once we have determined that conscious experiences are identical with particular brain states there is no obligation for us to explain why they are identical. To make this point he imagines the Mark Twain and and the Samuel Clemens historical appreciation societies who coincidentally hold their conventions at the same

[37] He derives the phrase 'antipathetic fallacy' from John Ruskin's pathetic fallacy, which concerns the erroneous attribution of human feelings to objects in nature in virtue of the figurative language of poetry, e.g., 'the deep and gloomy wood' (see Papineau 1993, 116).

venue. When their members meet they come to realise that their societies are dedicated to the same man. If they discovered that Mark Twain is Samuel Clemens in this way we would not be inclined to ask why they are the same person. As Papineau explains: "If they were [the same person], they were, and there's an end on it" (*ibid*, 121). Likewise, he argues, if consciousness turns out to be identical with some complex physical property *A* it makes no sense to ask "*why* consciousness is always present when physical property *A* is" (*ibid*).

However, these cases are importantly disanalogous. In the case of the appreciation societies, 'Mark Twain' and 'Samuel Clemens' unequivocally refer to individuals; we have clear criteria of individuation with respect to them. Thus we are able to pick out satisfactorily the person, or at least some historical facts about him, to which the names refer. But in the case of identifying a type of quale with some physical property there is no criterion of individuation for qualia *prior* to the determination of the identity. That is, until we assume that qualia are brain states we cannot identify particular qualia. What is needed in order to *discover* that qualia are in fact identical with brain states, in the same sense as the societies' members discover that Mark Twain and Samuel Clemens were the same person, is an independent means of identifying particular qualia. Without such an independent criterion of individuation it is question-begging to assert that we can discover that qualia are identical with brain states. So, to reiterate, a criterion of individuation for qualia in terms of physical properties presupposes their identity with physical properties. To assert that person S realises a red quale, for example, if and only if S's brain has properties $P_1,\ldots, P_n$, is to identify the red quale with the brain state that has these properties. In other words, we cannot determine that a red quale is identical

with some brain state without first presupposing this by a physical criterion of individuation for qualia. No such question-begging occurs when the members of the appreciation societies conclude that Mark Twain is Samuel Clemens.

It might be replied, however, that while we may not be able to show directly how particular qualia are identical with some brain state there is plenty of evidence to support the empirical hypothesis that qualia are identical with brain states quite generally. To this extent Flanagan points to the ways in which qualia can plausibly be individuated by physiological criteria. He cites Clyde Hardin's research in this regard which concerns colour perception and the possibility of spectrum inversion. It is widely assumed that we cannot rule out spectrum inversion by appealing to our physiologies and other physical facts about us. Hardin, however, argues that we can pin down relations between colour qualia and the neurophysiological states said to realise them, and therefore resist concluding that spectrum inversion is possible. To do this requires being able to identify a particular colour quale. Let us look at Hardin's argument.

3.4.1 Evidence for the Identity Theory?

Hardin identifies a commonly held view, namely, that qualia bear no *relations* to physical properties. A classic expression of this view that Hardin quotes is by Leibniz in his *Monadology* (see Leibniz 1998, 270). There Leibniz imagines scaling up the brain so that we could walk into it as we can into a mill. Assuming that the physical processes of the brain are inherently no different to the workings of a mill, Leibniz notes that observing in this way the workings of the brain would obviously not lead us to an understanding of how it is that the brain perceives anything. Nothing about the neurophysiology, therefore,

can tell us anything about perception, or more precisely, why it is that each thing seems to us a particular way rather than some other way. This view, Hardin thinks, is evidenced in Joseph Levine's well-known discussion of the problem of the explanatory gap (Levine 1983). There Levine denies that we can individuate colour qualia in terms of physical properties. Call the physiological account of a person's seeing red "**R**" and her seeing green "**G**". Levine contends that "it seems just as easy to imagine **G** as to imagine **R** underlying the qualitative experience that is in fact associated with **R**. The reverse, of course, also seems imaginable" (1983, 357-58). Thus, according to Levine, there are no necessary connections between colour qualia and the physiological processes said to underlie them. Whatever physical account **R** of a person's experience of seeing red it is always possible to suppose that this person is in fact experiencing green, i.e., realising a green quale.

In reply Hardin argues that given a detailed enough version of **R** this possibility can be ruled out. Nonetheless, he admits to offering no proof against the possibility. That is because of our lack of detailed knowledge about brain processes involved in colour experience. In this sense his reply is speculative. But, he thinks that it is very likely that when we do gain such knowledge we shall be able to rule out the possibility. Hardin suggests that input from our different sensory modes are to some degree compared. As he notes, "there is a considerable body of reliable cross-modal psychological connections documented in the experimental literature" (1987, 290). Assuming this is so **R** must include a description of how the perception of red is related to other sensory modes. We already have evidence of such relations. For example, colours are given somatic descriptions in terms of their advancing or receding, or in terms of being warm or cold.

Hardin contends that this evidence strongly suggests that there is some part of the visual cortex which acts as a central processing unit that compares inputs from the sensory modes with our colour perceptions. Accordingly, the accounts **R** and **G** will inevitably differ in detail.[38] And for this reason we can ultimately rule out the possibility that a person in physical states underlying the experiencing of red could, instead, be experiencing green.[39]

Like Flanagan Hardin thinks that we can understand qualia relationally. As we have seen, he argues that our colour experiences are complexly structured such that it is very difficult to imagine that when a person is perceiving red, say, her experience could be

---

[38] In general, therefore, Hardin claims that which colours we have the capacity to perceive is determined by the physical structure of our perceptual apparatus. Our colour perceptions, he explains, are coded on two chromatic channels. Each of these channels has a neutral, or base, level of excitation. In this neutral state no colour is perceived. But if these channels realise levels of excitation above their base rates, then we perceive either yellow or red, depending on which channel it is. And when they are excited below the base rates we perceive blue or green respectively. Thus, the two chromatic channels have either yellow-blue or red-green opponent colours. This fact tells us we should not be able to perceive a combination of these opponent colours since, of course, a channel cannot be both excited above and below its base rate simultaneously. Since this is indeed the case it is reasonable to conclude that how we perceive colours is determined by the underlying physiological processes. This again seems to suggest that we can individuate the qualia a person realises in terms of her physiological states, or properties more generally.

[39] Hardin considers two replies to his denial of colour qualia inversion. First, a person's colour qualia might be epiphenomenal; so she realises a green quale looking at a red object, for example, but her behaviour is independent of this fact – she behaves *as if* the colour of a ripe tomato elicits warm feelings, so to speak, despite its eliciting cool feelings. Second, it might be possible that the inversion includes comparisons with other sensory modes, so that the person not only realises a green quale looking at a red object, it also seems to her to be a warm, advancing colour. The first possibility Hardin finds outlandish (1988, 138). It requires us to suppose either that a person can believe that red seems cool to her but is unable to express this belief, or that she expresses a belief that red seems warm to her but experiences it as cool. With respect to the second possibility, Hardin points out that it is difficult to imagine how a person could realise a green quale and yet the colour seems warm to her. Such a quale would effectively be a residue of green, as he puts it, that "would seem to correspond to nothing in experience or imagination" (1987, 292).

thought of as being of the colour green instead. Consequently, he rejects Levine's claim that whatever physiological account of seeing a particular colour, C, we offer, it is always possible to imagine that the person has this experience while realising a different colour quale to C, e.g., the quale of its complementary colour in the case of spectrum inversion. However, Hardin's denial of Levine's claim *presupposes* that such a physiological account of C identifies C's quale. More precisely, in his discussion of Levine's position Hardin conflates experiencing a colour and realising a particular colour quale. He treats these as the same thing. He states that "Levine asks us to decide whether one could be in physical state **R** and yet have an experience of green, without letting us know anything about the particulars of **R**" (1987, 286). But in fact Levine asks us to decide whether one could be in physical state **R** and yet realise a green quale, i.e., that the *qualitative character* of the experience be green rather than red. Levine is not suggesting that in physical state **R** a person might be having a green experience as such. This is a subtle difference but an extremely important one. Levine assumes that experiencing a colour C as describable in physical terms and realising a C-quale are distinct. Hardin, on the other hand, takes these to be the same thing. So, like Papineau, Hardin takes qualia to be individuable on the assumption that to have an experience of C is identical with realising a C-quale. Again this is circular; it is simply to presuppose that qualia are identical with brain states.

It is this sort of question-begging that Dennett points to when he insists that the reasons a neurophysiologist might give for preferring one hypothesis concerning qualia over others only follow from her appropriating the term "qualia" to her own theoretical ends. Here, I think that Dennett is pointing out that in order to endorse the one hypothesis

the neurophysiologist must assume a physical criterion of individuation for qualia. And that is to think of qualia as relational properties which again, as Dennett notes, "takes us too far away from our original idea of qualia as properties with which we have a particularly intimate acquaintance" (1988, 60). If we think of qualia as relational properties in this way, it is very difficult to understand how they can have a subjective aspect. Flanagan's reply to this difficulty is to suggest that qualia are subjective in the sense that they can be characterised as the ways things seem to us. Characterising qualia in this way, according to Flanagan, allows us to pin down "the phenomenological features of mind so that we can check for relations among the phenomenological, psychological, and neurological levels" (1992, 61). However, 'the ways things seem to us' describes an *essentially* subjective aspect of experience, my pain quale with respect to a toothache, for example, is that property of my pain experience intrinsic to me, i.e, the way the toothache seems to me. Understood in *this* way qualia are not relational properties. Flanagan claims that we can conceive of qualia more broadly to include their relations to natural properties. But, how is this possible unless we have some way of understanding qualia *as subjective phenomena* relationally? If there is a conceptual link between qualia so understood and physiological properties, that is, a way of understanding the subjective aspect of qualia in physiological terms, then qualia could straightforwardly be thought of as relational properties. But no such link exists.

3.5 Summary

I started, in section 1, by laying out a conception of qualia, an account of how we think of them, that is as uncontroversial as possible. Crucial to our understanding of qualia, I

argued, is that they are perspectival in nature. This is the essential insight about qualia that originally led Nagel to anticipate the problem of consciousness, which he expressed in terms of their dependence on a particular point of view. The peculiarity of qualia in this respect is that we are each directly acquainted with them, and as such we do not have to posit them in our scientific and common-sense theories about the world as we posit ordinary properties such as blueness, being a whale, being one metre high, and so on. The difference is that we apprehend these ordinary properties, either directly or indirectly by our senses, whereas qualia are those properties in virtue of which we apprehend these ordinary properties. That is what is peculiar about qualia. In understanding qualia thus, I then argued, it is best to think of them as constitutive of the subject of experience, i.e., the phenomenal subject. The idea is that the term 'qualia' describes the manifestation of the phenomenal subject. Qualia exist as the phenomenal subject.

Qualia thought of this way must be, however, inapprehensible even by the subject. This follows from the fact that as constitutive of the subject it is impossible that the subject can apprehend them. The inapprehensibility of qualia, I argued, implies that they are unindividuable. That is, there can be no criterion by which we can judge if one quale is distinct from or identical with some other quale. In section 3 I considered Dennett's arguments aimed at eliminating qualia. What is interesting about his arguments is that he likewise claims that qualia are unindividuable, albeit towards a different end since I accept the existence of qualia *pace* Dennnett. And following Dennett I discussed aspects of Wittgenstein's private language argument relevant to the unindividuability of qualia. Wittgenstein's views on the issue will become central to discussions in chapter 4 especially. In his arguments Dennett points out that insofar as we can think of qualia they

cannot be relational properties in any sense, and so there is no hope of understanding them in behavioural or physiological terms.

The conclusion that qualia are unindividuable is compatible with how we find qualia to be, that is, it agrees with our inability to apprehend other people's qualia or to understand even what this could amount to. However, their unindividuability does make it impossible to identify them as brain states, or as anything else for that matter. This would preclude identifying qualia as physical properties, and *a fortiori* it would preclude identifying experiences as brain states to the extent that our experiences are identified by their qualia – a person's experience is thought to be one of pain, for example, only if the experience has a pain-quale, i.e., the right kind of phenomenological quality. Accordingly, identity theorists must assume qualia are individuable. As I indeed point out in section 4, the identity theorist argues that qualia are straightforwardly individuable because they are identical with certain types of physical properties. In this section I considered Flanagan's objections to Dennett's claim that qualia are unindividuable, and then I broadened the discussion to include arguments against the general claim made by Papineau and Hardin. With respect to all three philosophers – Flanagan, Papineau and Hardin – I argued that their reasoning is question-begging. For example, Papineau contends that there is overwhelming empirical evidence suggesting that qualia are identical with properties of the brain, or at least to deny that they are identical is too problematic. But even to assume that qualia can be identical with certain physical properties entails already assuming that they are individuable. None of these philosophers, I argue, succeeds in showing that qualia are individuable.

But, as I noted in the introduction to this chapter, if qualia are unindividuable then they are not naturalisable. Qualia cannot be subsumed under our scientific theories if we are unable to determine when two or more instances of a quale fall under the same type or under distinct quale-types. This conclusion seems to contradict my conclusion in chapter 2 that there are no *a priori* reasons to suppose that consciousness is not naturalisable. Again, I shall argue in the next chapter that this contradiction is only apparent.

Chapter 4

Saving the Phenomenological

I have argued that qualia are unindividuable and hence cannot be naturalised, but that consciousness is nonetheless naturalisable. Thus consciousness, as a natural phenomenon, can be understood in terms of behaviour and physiology, and qualia play no direct role in this understanding. This conclusion concurs with Dennett's view that consciousness is analysable in heterophenomenological terms, which suggests that the concept of qualia is mistaken or empty so that talk of qualia is about nothing. Indeed, Dennett argues that there are no such properties as qualia. I agree that consciousness can be naturalised in heterophenomenological terms, but I resist the conclusion that qualia do not exist. Here we struggle with words, but denying the existence of qualia makes it difficult, if not impossible, to understand ourselves as experiencing anything, in contrast to being something like a chair, namely, something that there is nothing it is like to be. Thinking of qualia concerns the attempt to account for the subjective aspect of consciousness that distinguishes it from other phenomena. Thus, to the extent that Dennett eliminates the subjective aspect of experience he is often accused of changing the subject. That is to say, for many what is essential to consciousness is its phenomenology, i.e., those features that

seem apprehensible only from the first-person viewpoint, features that a heterophenomenological analysis misses out. Therefore, the phenomenon Dennett takes to be naturalisable is not consciousness as such. Below, I aim to reconcile the tension between maintaining that the concept of qualia plays no role in naturalising consciousness and that qualia are real properties.

After taking stock, in section 4.1, of the main conclusions reached so far, I point to the principal aim of the chapter, namely, to reconcile the two following apparently conflicting claims: (1) Qualia exist and (2) Qualia are unnaturalisable. In section 4.2 I again look at how Dennett argues for the unindividuabilty of qualia, and explain why we do not have to accept his conclusion that their unindividuability requires us to deny their existence. In section 4.2.1 I argue that *prima facie* we seem able to think of someone being conscious, as measured by behaviour and physiology, independently of being conscious phenomenologically understood in terms of realising qualia, but these two apparently distinctive ways of thinking of consciousness are mutually dependent; that is to say, they are two aspects of a *single* concept in the sense that the one way of understanding consciousness presupposes the other. This mutual dependence derives from the fact that we cannot acquire the idea of being conscious in the one sense without the idea of being conscious in the other. Consequently, although consciousness is only naturalisable in terms of a naturalistic understanding of consciousness, i.e., in terms of behaviour and physiology, this fact does not make our phenomenological understanding of consciousness irrelevant. The very fact that we have a naturalistic understanding of consciousness requires this other kind. Then in section 4.2.2 I argue the concept of consciousness understood in terms of this mutual dependence shows how the zombie

hypothesis is only superficially plausible. And in section 4.2.3 I consider an objection to this understanding of the concept of consciousness in terms of the problem of the explanatory gap, showing how this objection can be met. Finally in section 4.3 I briefly consider how the unindividuability of qualia might threaten to make talk about qualia meaningless, a worry which again is deflected.

4.1 Making Sense of Consciousness

In the first chapter I argued that it is a good policy to adopt naturalism, that is, an approach or attitude based on treating philosophy as continuous with science, justified principally by the fact that there is no higher authority by which to determine how the world is than our senses. Any hypothesis that is putatively about the world that contradicts or is in conflict with what our senses tell us, roughly speaking, must either be rejected or adjusted to explain away this conflict. By this measure, no metaphysical claim based on intuition can falsify any scientific hypothesis held true in virtue of its consistency with the rest of our best scientific theories and its agreement with our experience. The necessity of empirical falsifiability in this sense means that metaphysical claims cannot stand above scientific scrutiny. That is to say, our scientific theories, and practices more generally, cannot be undermined by metaphysical considerations because such considerations cannot appeal to a higher authority.

Nothing stops us from assuming philosophical hypotheses can be held true independently of any ultimate empirical confirmation. But I have argued that a non-naturalistic attitude, based on assuming that such philosophical hypotheses can undermine our scientific theories, tends to lead to extravagant or perverse metaphysical

claims. The example we looked at is Chalmers' suggestion that perhaps consciousness is ubiquitous, rather than a chauvinistic biological property. Because this hypothesis is logically independent of the rest of science and untestable it cannot be shown to be false. Nor can it be shown to be true in any robust sense, specifically, in terms of being compatible with how we find the world to be according to our senses. Chalmers argues that this hypothesis can be held true to the extent that it is part of a plausible theory of consciousness. The problem with trying to measure the truth of a theory by its plausibility *alone* is that it is an intuitive measure rather than an objective one. Truth concerns how the world is, not how we intuitively think it must be. Underlying this observation are the facts about how we conceive of natural phenomena.

To understand a phenomenon as natural is to suppose that we think of it according to how well it fits with our theories as confirmed by observation. To borrow Carruthers' term mentioned earlier (see section 2.4.1), natural phenomena or properties are understood as being thickly individuable. That is, a single natural property can be individuated, or picked out, by various concepts. There is no one-to-one correspondence between properties and our concepts of them such that concepts determine properties. To assume otherwise, Carruthers notes, leads us to endorse a highly implausible platonism.[40] However, it is important to note that Carruthers embraces a full-blooded realism about natural properties. He tells us that

---

[40] Briefly, the claim is that every valid concept must correspond with some mind-independent property, but concepts are mind-*dependent*. Accordingly, our minds would have to be capable of grasping so-called mind-independent properties. The only way to make sense of this possibility is to think of properties as inhabiting a sort of abstract platonic realm accessible to our minds (Carruthers 2001, 35-36). The problems with such platonism are well-known, and there is no need to cite them here; in particular, it is antithetical to the idea of what natural properties are.

> we should believe, both that there are real properties belonging to the natural world, and that which properties there are in the world is an open question, which cannot be read directly off the set of concepts which we happen to employ (*ibid*, 37)

Here and elsewhere Carruthers suggests that what properties there are in the world is independent of our concepts in some absolute sense, as when he describes this as an 'open question'. Carruthers, therefore, thinks of properties as somehow fully entities in their own right. I would qualify his assertion above by insisting that we cannot think of properties as wholly or absolutely independent of our concepts. Picking out a property requires possessing a corresponding concept. Fortunately, I think we can subscribe to a more modest, less inflated, view of properties than the one embraced by Carruthers. We can instead think of properties in terms of predicates, such that an object is said to have a property if some predicate is true of that object; and in this sense properties can be said to exist. So, for example, we say that there is a property of being furry if it is true, for example, that some cat is furry.

Now, Carruthers is drawn to his property realism by the thought that how we think of an object, i.e., in terms of what properties it has, cannot determine how it really is in this sense. By this measure, we cannot ever assume a one-to-one correspondence between properties and concepts – it is always possible that we have more than one way of conceiving of some property. Hence, the relationship between properties and concepts is one-to-many rather than one-to-one. However, we can understand *talk* of one-to-many correspondence between properties and concepts other than in the way Carruthers

understands it. Instead we can understand *predicate terms* as being semantically equivalent. So, rather than talk of an object having the same property differently conceived, we can talk of two or more predicate terms, e.g., *F* and *G*, having the same meaning so that *F* and *G* attribute the same property to the object. Accordingly, there would be a one-to-many relationship between properties, thought of as concept predicates, and predicate terms. So, to use a previous example, rather than thinking of there being two concepts, R and S, that seem to refer to distinct properties, i.e., diseases, but which are later discovered to refer to a single property, we can think of there being two predicate terms, *R* and *S*, that seem to pick out distinct concept-predicates, but which are later judged to have the same meaning and hence can be said to pick out the same property (see section 2.4.1).

The view of properties I have outlined above, this other way of thinking of properties, might reasonably be called conceptualist. However, I am reluctant to describe it as such. That is because conceptualism about properties is usually defined as the view that properties are mind-dependent. Thus, it is understood to be the rejection of the idea that properties exist in their own right. I do not think that this is a helpful way of viewing things. Talk of properties in their own right is an attempt to elucidate the thought that how the world is does not depend on how we think of it. But it is implausible to think that we can understand this talk as implying that independently existing properties are identifiable as such. This is what I reject in Carruthers' view. Carruthers holds that science aims at, and could eventually succeed in, picking out properties in their own right (*ibid*, 38). This is to presume a dichotomy between mind-dependence and mind-independence, so to speak. Thinking in terms of this dichotomy is unhelpful.

Science does not reveal properties in their own right. We are tempted to suppose that science provides an ever fuller or more comprehensive *picture* of the world in terms of properties. This picture of the world might be likened to a written portrait of a person, describing in increasing detail her character, looks, etc. such that it corresponds ever more closely to the real, i.e., mind-independent, person. However, with respect to science there is nothing analogous to this corresponding 'given' person, nothing we can hold up this so-called picture of the world against. We do not have to think of science as a picturing *of* the world, rather we can think of it as an activity *in* the world. Does the truth of our best physical theories, for example, imply that some things *really* are negatively charged? Yes, but it is nonetheless a misleading question. What our best physical theories show is that we must think of certain things as being negatively charged – it fits with our observations in general, and it enables us do things in the world. Importantly, this is not usefulness in the sense implied by pragmatism, that is, concerned simply with getting things done; rather it concerns the hard and fast notion of usefulness for science, based on the demand that our scientific theories are predictively successful, so that if a theory fails to be it is our duty to alter it.[41] Carruthers, on the other hand, seemingly bewitched by this notion of science as the picturing of the world, betrays a non-naturalistic attitude by presupposing that there is a God's-eye-view that science aims to reach, a viewpoint on the world as it is 'independent of us' in some absolute sense. No such viewpoint exists. The view I am urging is realist insofar as the properties posited by our best scientific theories truly exist, i.e., are real, on the understanding that what we believe to be real is ultimately

[41] Here, I have in mind William James's notion of truth, as expressed in various passages in his *Pragmatism*, 1907, as it relates to his so-called pragmatic method.

putative – we must always be prepared to give up our belief in the existence of properties if adjustments in our theories demand that we do. This is not some second-rate realism, for as Quine remarks: "Nor let us look down on the standpoint of the [scientific] theory as make-believe; for we can never do better than occupy the standpoint of some theory or other, the best we can muster at the time" (1960, 22).

With respect to consciousness we have seen that Dennett argues that 'qualia' represents a 'confused intuitive concept'. In this regard he claims that the philosophical concept of qualia is based on a loose way of talking in terms of 'how things seem' to each of us. Indeed, it seems unexceptional to say, for example, that how green things seem to me is how they seem to you. In the everyday context this way of talking leads to no confusion inasmuch as little depends on its being true or false, since such assertions are not expected to be tested in any rigorous manner. And that is a good thing given that qualia, understood as the phenomenological qualities of our experiences, are unindividuable. That is, there is no criterion by which we can ever determine whether some quale is in fact identical with some reportedly distinct quale or not from the third-person viewpoint. It is not possible, for example, for me to know that how a toothache feels to me is the same or distinct from how one feels to you. To the extent that knowledge requires providing *independent* evidence, there seems to be no way of knowing such things. We have no means of identifying, or picking out, individual qualia from one another. However, unless an entity is individuable in this sense there is no way to quantify over it since, to cite Quine's maxim, we can have 'no entity without identity'. This implies that qualia cannot be naturalised, since to naturalise an entity, such as a property, requires being able to quantify over it in order to bring it under a theory.

Very broadly there have been two responses to the unindividuability of qualia. One response has been to conclude that because qualia are self-evidently real, i.e., exist, the only way to explain their unindividuability is to suppose that they are not physical properties, since they are accordingly not apprehensible from the third-person point of view. This roughly characterises Jackson's original position (1982), for example. The second response has been to assume, with the non-physicalists, that qualia are real, but to conclude that they are unindividuable because we cannot properly conceive of consciousness, and *a fortiori* of qualia – it is at least beyond our present conceptual capacities, and maybe forever so. Accordingly, qualia are in principle individuable but we could never grasp any criterion by which to do this. This describes the kind of mysterian response offered by McGinn (1989/91), and with which Nagel is sympathetic.

By now it should be clear that these responses face serious difficulties, and consequently neither of them is attractive. But as I argued earlier, to insist instead that we can individuate qualia fails. As we have seen, this is the basic position of identity theorists like Flanagan and Papineau, who argue that qualia are physical and hence individuable, at least in principle. The problem is that they provide no independent support for the claim that qualia are physical. Although Flanagan, for example, argues that while we do not understand how qualia, thought of as physical properties, have a phenomenological aspect to them, we can at least come to understand more clearly how phenomenology arises in the brain as our neuroscience advances (Flanagan 1992, 35-60). The trouble with this claim is that *in essence* we understand qualia phenomenologically, and no amount of detail about the underlying mechanisms in the brain can explain qualia so understood.

The unindividuability of qualia, then, seems to entail an undesirable dichotomy since neither response is acceptable. I think, nonetheless, that there is an element in the responses that is right. And so what I want to explore is the possibility of conserving what is right about each of them and weaving this into a coherent position. There are two fundamental insights I think that we should accept:

(1) Qualia exist; there is a phenomenological aspect to experience.

(2) Qualia, as properties of experience, are not naturalisable, as they are unindividuable and hence cannot fall under any naturalistic theory.

However, one might argue, as Dennett does, that qualia are not naturalisable because the concept of qualia is incoherent. And this, Dennett contends, shows that there are no such properties as qualia. But such eliminativism about qualia seems very difficult to accept. As Levine, for example, explains somewhat rhetorically:

> [W]hat could be more obvious than the fact that we have conscious sensory experiences? How could you deny that there is something that it's like to see red, smell a rose, or feel pain? What possible illusions could we be suffering from in thinking that these are all genuine properties of experience? (2001, 131)

The idea that qualia do not exist does appear absurd. But does the unindividuability of qualia, specifically, imply this? The assumption that qualia are unindividuable concerns the *concept* of qualia. It is to assert that insofar as we think of consciousness in terms of the properties of experience, i.e., phenomenologically, these properties do not have

identifiable instances. But that is not to suppose that we cannot conceive of qualia generally; that is, it can still be assumed that, for example, there is something that it is like to see the colours yellow and blue for most of us. What is denied is that we can talk about individual qualia. In other words, while we can say that there is something that it is like to see yellow, blue, red etc. in general, talk of how it is to see red *as opposed to* yellow, say, goes nowhere.[42] Therefore, to assert that qualia are unindividuable does not seem to force us to conclude that there is nothing that it is like to smell a rose and so on. Here we are reminded of Wittgenstein's remark concerning his beetle-in-a-box scenario, cited earlier in section 3.3, that while the content of the box "is not a *something*, it is not a *nothing* either!" (1958, 304). In other words, while the concept of qualia does not concern particular property instances it is not about nothing, i.e., qualia do exist.

4.2 Retaining Qualia

How exactly might one suppose that the unindividuability of qualia implies that they do not exist? One might reason that if it is impossible to determine that one quale is in fact distinct from another, then any statement about qualia, e.g., 'I am having a red quale looking at this tomato', is unfalsifiable in principle. This fact suggests that it must be denied, in the above example say, that I am *really* having a red quale since if I were so it

[42] That said, this is not to deny that we can make sense of the notion of distinguishing between qualia *simplicter*. Whatever it is like for someone to see red and to see yellow, we can straightforwardly assume the person distinguishes between these qualia, as evidenced behaviourally. Indeed, as Kim remarks, "the intrinsic qualities associated with qualia are, or may be, undetectable, but differences and similarities between qualia, within a single individual, are behaviorally detectable, and this opens a way for their behavioral functionalization" (2005, 172).

would be confirmable. It is this essential line of argument that Dennett uses to deny that qualia exist.

Dennett is sometimes interpreted as appealing to verificationist principles in his argument for the claim that qualia do not exist. This is how William Seager, for example, interprets Dennett. Seager helpfully summarises verificationism in general as "the doctrine affirming that where one can't tell the truth about something, there is no truth to tell" (2001, 85). This describes the verificationist principle of *truth*, otherwise expressed as 'a statement is true if and only if it is verifiable'. Dennett's argument also would seem to appeal to the verificationist principle of *meaning*. This principle, most often referred to as the principle of verifiability, states that a statement is meaningful if and only if there is some method of verifying it. As Dennett makes clear at the beginning of his 'Quining Qualia' (1986), his aim is to show that the concept of qualia is incoherent. In this sense his general strategy is to point out that statements concerning qualia are unverifiable, and this fact suggests that the statements are in an important sense meaningless, and hence there is nothing that such statements are about.

Consider a statement that involves a pseudo-concept in this way: 'God is perfection'. This statement could be said to be meaningless because we have no method by which to verify it empirically. To verify this claim we need to identify God, and this we cannot do since the concept of God bears no relations to anything observable. Thus, appealing to the verificationist principle of truth, if this statement is unverifiable then it cannot be true. Therefore, the term 'God' fails to refer, hence God does not exist. Dennett is taken to argue that statements concerning qualia are meaningless for the same kind of reasons. But insofar as Dennett can be interpreted as arguing that qualia are unindividuable – a

position, I have argued, that can be very plausibly attributed to him – he does not need to appeal directly to these verificationist principles.

Specifically Dennett argues that the concept of qualia is confused such that the term is unusable – any statement employing the term 'quale' is unverifiable. He observes that the concept originates from a pretheoretical notion (1986, 227). He describes 'qualia' as "an unfamiliar term for something that could not be more familiar to each of us: the *ways things seem to us*" (*ibid*, 226). As he remarks, this original notion is "so thoroughly confused that even if we undertook to salvage some "lowest common denominator" from the theoretician's proposal...it would be tactically obtuse – not to say Pickwickian – to cling to the term" (*ibid*, 227). He elucidates how this everyday notion relates to that of qualia using examples such as the way milk tastes to someone is a gustatory quale (*ibid*). Dennett points out that the trouble with the concept, to repeat what was written in an earlier chapter, is that qualia are *thought of* as essentially private and intrinsic properties so that reports concerning someone's qualia are empirically unverifiable. In other words, qualia would have to be properties whose instances cannot be independently identified; that is to say, qualia would have to be unindividuable. It is impossible to determine whether a quale is distinct from or identical with another quale from a third-person, i.e., objective, viewpoint. By this measure, the extension of the concept of qualia is indeterminable so that the set of actual instances of qualia cannot ever be determined. Hence the concept is unusable in the sense outlined above, namely, any statement employing the term is unverifiable.

In this way his argument does not appeal directly to verificationist principles. We do not conclude that the concept is incoherent *because* judgments concerning it are

unverifiable and hence cannot be true; rather, we can take Dennett to be arguing that the concept is incoherent and this is *evidenced* by the fact that judgments concerning them are unverifiable. In essence his argument would therefore amount to the following. The concept of qualia, based on our everyday notion of how things seem to us, is confused and incoherent. How we think of qualia entails their being unindividuable. Their unindividuability implies that the concept's extension is indeterminable. This is evidenced by the unverifiability of judgments concerning them. Consequently, we are better off giving up on the concept; or, as he puts it, it is "[f]ar better, tactically, to declare that there simply are no qualia at all" (*ibid*).

However, I think Dennett is too pessimistic. I agree that ordinarily if we were to think of some property as unindividuable, then the concept would be unusable, and hence we would be best off passing it over. But, it is precisely the unindividuability of qualia that *defines* the concept, and so rather than abandoning the concept we should aim to understand why they are unindividuable. Consider the following hypothetical example. Imagine someone observes an astronomical phenomenon that he calls a 'mysterion'. A mysterion, he proposes, does not necessarily have a single location, so that if we ostensibly observe two mysterions simultaneously at different points in the sky we are unable to determine whether they are the same mysterion or distinct ones. Mysterions, therefore, would qualify as unindividuable phenomena. Now, there seem to be two ways in which mysterions could be thought of as unindividuable. First, we might assume that there is a fact of the matter as to whether two such observations are of the same mysterion or distinct ones, but we cannot know which is the case. This is to think of mysterions as being *epistemically unindividuable*. Second, we might think that there is no

fact of the matter vis-a-vis such observations. This is somewhat analogous to the indeterminacy of the position or direction of a subatomic particle in accordance with Heisenberg's Uncertainty Principle. That is, the position of a subatomic particle is not determinate *until* it is observed; except that in this case not even observing mysterions would determine their identities. Let us call this *ontological unindividuability*.

The idea of ontological unindividuability is plainly incoherent. It amounts to assuming that it is undetermined whether a so-called mysterion is identical with itself or not, which is nonsense. But what can we make of the idea of the epistemic unindividuability of mysterions? Let us call the observed mysterions at two points in the sky '*A*' and '*B*'. The thought is that it is either the case that $A = B$ or that $A \neq B$ but we can never know, i.e., there is no criterion by which to determine which is the case. If this is so, then clearly the concept of mysterions is unusable logically, that is, no judgment concerning them would ever be verifiable and hence statements concerning them are not truth-evaluable. Accordingly, we are best off to give up on the concept of mysterions altogether. But as an *epistemic* limitation it might turn out that there are reasons why mysterions are unindividuable. This is at least a hypothetical possibility. We could still maintain that they exist so long as we could offer a satisfying account of why we cannot *know* when two observed mysterions are the same individual or distinct individuals. And the same holds for qualia, that is, if we can explain why qualia are epistemically unindividuable, then we do not have to conclude that the concept is incoherent and should consequently be given up, as Dennett recommends. Indeed, later I shall offer the outline of such an account.

4.2.1 Two Ways of Understanding Consciousness

Earlier I argued that insofar as we attribute consciousness to others we do so *strictly* in terms of another person's or creature's behaviour and physiology. Whether *others* have an inner life, i.e., whether they are phenomenologically conscious, plays no direct role in judging them to be conscious as such. When we say, therefore, that someone is conscious we are attributing to them such-and-such behaviour and physiological characteristics. Only by thinking of consciousness from this third-person viewpoint can we get the concept of consciousness off the ground at all. If, on the other hand, we thought of consciousness strictly from the first-person viewpoint, then my judgment that I am conscious would be distinct from my judgment that you are conscious, i.e, they would be justified in wholly different ways. Let us call this third-personal understanding of consciousness 'consciousness$_N$'. And let us call our phenomenological first-personal understanding of consciousness 'consciousness$_p$'.

Now, consciousness as we understand it phenomenologically, i.e., from a first-person viewpoint, appears to be a property distinct from the one we conceive of naturalistically, i.e., from the third-person viewpoint. Here it is not being suggested that because we can think of consciousness$_P$ and consciousness$_N$ separately they must be distinct properties. I am not reviving the conceivability argument discussed in chapter 1. Rather, it is being pointed out that to the extent that these ways of thinking of consciousness are independent of each other we seem unable to rule out the *possibility* that they concern distinct properties. This possibility is starkly illustrated by the zombie hypothesis. A zombie is essentially thought of as a being that is conscious$_N$, that is, it exhibits all the right kinds of behaviour and possesses the relevant physiological characteristics so that

we can judge it to be conscious_N, but it is not conscious_P, that is, it has no phenomenology. Such a being is logically possible, however odd it strikes us, if nothing about how we think of consciousness_P implies consciousness_N.

So, unless it can be shown how these two seemingly distinct ways of thinking of consciousness are dependent on one another, i.e., they are effectively aspects of the same concept, we cannot rule out the possibility that these 'concepts' concern different properties. If they do concern different properties, then the claim that consciousness is naturalisable is far less interesting since it amounts to the claim that consciousness_N only is naturalisable. Further, the fact that I have argued that qualia are unnaturalisable – and therefore consciousness_P is unnaturalisable since it is understood in terms of qualia – seems to suggest that these two aspects of consciousness do indeed concern different properties. Only if we can understand consciousness_P and consciousness_N as aspects of a single concept is it at all plausible to claim that consciousness is naturalisable. Below, I argue that these two ways of understanding consciousness are mutually dependent on each other, and their mutual dependence implies that they are effectively aspects of the same concept. That is because the idea requires us to think of consciousness_P and consciousness_N as distinct and yet we cannot think of the one independently of the other.[43]

[43] This same basic point has been made by P.F.Strawson (1959). Essentially Strawson argues that we can ascribe conscious states to ourselves only if we can ascribe them to others, and we do this in virtue of certain bodily states we observe others to have. If, on the other hand, we did not for example ascribe pain to others according to certain behavioural and physiological states we observe, then it could not be explained how we ascribe pain to ourselves. Or as Strawson puts it, "[t]he condition of reckoning oneself as a subject of such predicates [i.e., predicates concerning conscious states] is that one should also reckon others as subjects of such predicates" (1959, 100). And one can only

Let us see how consciousness$_P$ and consciousness$_N$ are mutually dependent. In general the mutual dependence of two concepts concerns our being unable to possess the one concept without possessing the other. In this respect the concepts of being coloured and of being red are dependent – a dependence understood in terms of the determinate/ determinable relation between these properties in this case. We cannot possess the concept of being red without possessing the concept of being coloured. In other words, we cannot acquire the concept of being red unless we possess, or acquire at the same time, the concept of being coloured. However, one might object that this dependence is not *mutual* in the sense that conversely one could possess the concept of being coloured without necessarily possessing the concept of being red – we can plausibly imagine, for example, a person who has not specifically acquired the concept of red but knows how to use other colour terms. But we can strengthen the example by imagining instead a superconcept consisting of the disjunction of colours, call it '*D*', such that '*D*x ≡ x is red or blue or green or yellow or...'. Here the concepts of being coloured and *D* are mutually dependent in the fullest sense, i.e., the dependence is symmetrical. In order to acquire the concept of *D* one must also acquire the concept of being coloured and vice versa. But, it still seems possible to imagine a person who has acquired the concept of *D*, i.e., who knows how to use all of the colour terms that comprise the disjunction but who does not know how to use the term 'coloured'. Faced with such a person we would have to

---

ascribe such predicates to others according to observable behavioural and physiological features. For Strawson this claim is used as a solution to the problem of other minds. How do we know that others have minds, i.e., are persons to use Strawson's terminology? In order to assume we each have minds it is necessary that we can ascribe mindedness to others first. In other words, the question how do we know others have minds is spurious given that it can only be asked when we think of others as minded to begin with.

conclude that she has tacitly acquired the concept of being coloured, and her inability to apply the term 'coloured' simply reflects the fact that her grasp of this concept is not explicit. Of course this example is hypothetical given that we do not imagine a person would ever use such a disjunctive concept. But it should be appreciated that the use of example is heuristic, the aim is to illustrate the idea of mutual dependence.

The mutual dependence of conciousness$_P$ and consciousness$_N$ is analogous to the way the concepts of *D* and being coloured are dependent, granted that conciousness$_P$ and consciousness$_N$ are standardly thought of as and seem to be distinctive concepts. Their inseparability in this respect can be shown by the following *reductio* argument. Assume instead that consciousness$_p$ and consciousness$_N$ constitute independent concepts, so that it is possible to possess the 'concept' of consciousness$_p$ without possessing the 'concept' of consciousness$_N$ and vice versa. Thus, we could each understand that others are conscious$_p$ like ourselves without concern for their behaviour and physiology. But then, how could I, under these conceptual circumstances, hold that a chair, for example, is not conscious$_p$? Short of discovering that we have a sixth sense that allows us to detect when others are conscious$_p$, it would be impossible for me to assume anyone else is conscious$_p$. Perhaps I could simply suppose that everything is conscious$_p$, i.e., adopt a full-blown panpsychism. But, by this measure the concept becomes too liberal, too general, for it to be of any use. Moreover, we saw with Chalmers' view that pansychism is very implausible, if not absurd.

Still, might it be possible to form the concept of consciousness$_p$ on the basis of evidence of others' behaviour and physiology without having to suppose that they really are conscious$_p$? I know that I am conscious$_p$ and all that is needed for me to share the

concept of $consciousness_p$ is to correlate my behaviour and physiology with my being $conscious_p$, and from the evidence of others having similar behaviour and physiology surmise that they understand what I mean by the term 'conscious'. However, by this account I do not ever suppose that others are conscious, i.e., hold it true. If the term 'consciousness' refers only to my phenomenology, then there is no sense in which others can know what the term refers to. Of course, we *do* suppose that others are conscious as we ourselves are. But to hold this as true we must assume that others are conscious on the basis of behaviour and physiology. That is to say, we must assume they are $conscious_N$. But that is what is denied in the reductio.

Conversely, imagine that we only possess the concept of $consciousness_N$. You would assert that others are $conscious_N$ on the basis of their behaviour and physiology, but this bears no relation to your being $conscious_p$. It requires you to dissociate behaviour and physiology from how you experience things, i.e., from the first-person viewpoint. So, for example, if you see someone bang her knee and then groan and grimace you would not associate this activity with how such an activity would feel to you. This is of course very difficult to imagine. You would effectively have to imagine yourself as if you were an entirely alien creature so that there is no sense in which people's behaviour and physiology concerns how things seem to you. What, accordingly, would attributing $consciousness_N$ to anyone amount to under these circumstances? This concept would be empty and consequently there would be no reason for it to originate at all. And, as I shall argue later, in a zombie world, i.e., a world populated entirely by beings behaviourally like ourselves, there would be no motivation, no reason, to talk about consciousness. The concept would be useless under such circumstances.

Thinking of consciousness$_N$ on its own essentially would amount to a behaviouristic conception of consciousness, i.e., construed not only in terms of a person's actions but also in terms of the movement or physical activities of the person generally, right down to neurological activities. Accordingly, it might be suggested that we can attribute to a person activities that can be thought of as conscious insofar as they fit specific dispositional states. For example, the sudden movement of an object in the periphery of someone's visual field (input) tends to lead to the person visually focusing on that object (output). But if such activities are understood solely in physical terms, i.e., as we understand them from the third-person viewpoint, then we need a reason to think of them as manifestations of consciousness, i.e., as conscious behaviour. In terms of their physical descriptions these activities are essentially no different from those activities we *in fact* judge not to count as conscious behaviour. In general, in order to possess a purely behaviouristic conception of consciousness we would need to be able to pick out those activities which are relevant to the concept. In other words, we would require a criterion by which to judge what does and does not qualify as *conscious* behaviour; and again from the third-person viewpoint no such criterion exists so that what we count as someone being conscious$_N$ becomes arbitrary.[44]

---

[44] Dennett, however, seems to suppose that consciousness can be thought of entirely in such behaviouristic terms. As noted earlier, he argues that we can understand consciousness wholly from the third-person viewpoint, in purely 'heterophenomenological terms' as he puts it. He asks us to imagine an advanced civilisation of Martians who visit Earth (2005, 25-56). These Martians might even be thought of as phenomenally unconscious, i.e., as so-called zombies, if, unlike Dennett, one thinks of such creatures as possible. Also they have a different set of perceptual apparatus to us. Over all, the idea is that these Martians' point of view, if they can be said to have one at all, is alien to us, that is, there is nothing we share with them in terms of how we are phenomenally conscious. In other words, the only way for them to come to

Therefore, it is very implausible to suppose that we can think of consciousness only in terms of either $consciousness_P$ or $consciousness_N$. And so we must conclude that $consciousness_P$ and $consciousness_N$ are mutually dependent – we cannot think of consciousness without taking these two perspectives into account.

4.2.2 The Zombie Hypothesis

If these perspectives on consciousness are inseparable in this way, however, it ought to be impossible for us to entertain the zombie hypothesis. That is, it ought to be the case that if we think of someone as $conscious_N$ we must also understand how she is $conscious_P$, i.e., her being $conscious_P$ should be transparent to us; and this is not so. In reply, we can only *superficially* think of someone's being $conscious_N$ separately from her being $conscious_P$. To understand how this is so again consider the example of the concepts of being red and being coloured. We can imagine someone thinking 'this snooker ball is red' without thinking 'this snooker ball is coloured'. This is a possibility in the sense that the person

---

understand consciousness is from the third-person viewpoint, i.e., from the perspective they do share with us. Dennett argues that this is possible. What the Martians do have in common with us is the ability to adopt what Dennett calls an *intentional stance*, that is, they are able to understand and talk about complex behavioural systems in terms of ascribing beliefs, desires etc., to such systems (see 1987, 13-35). This ability at least allows them to interpret our folk psychology. That said, it is clear from this description of the Martians that they themselves would possess no concept of consciousness. How, therefore, could they distinguish conscious from non-conscious behaviour from the start? Insofar as they would have developed their own folk psychology in virtue of being able to adopt the intentional stance, it would not include any allusions to how things seem to them and so on. Dennett essentially bypasses this difficulty by stipulating that the Martian scientists' aim is to understand *our* concept of consciousness, beginning with our folk theory of consciousness, from the third-person viewpoint. He writes that "[a]mong the phenomena that would be readily observable to these Martians would be all our public representations of consciousness" (2005, 26). What they take conscious behaviour to be, in other words, is determined by us given that we do understand consciousness in two ways at once, namely, from the first- and third-person viewpoints.

does not explicitly know how to use the term 'coloured'. Nonetheless, again, the person must have tacitly acquired the concept of being coloured if she understands the term 'red'. That said, the difference in the zombie hypothesis seems to be that even when someone explicitly knows how to use both terms, i.e., 'conscious_P' and 'conscious_N', she can still think of them separately. However, this claim is false. So long as someone has acquired the concept of consciousness_N she must at least tacitly have acquired the concept of consciousness_P. That is obvious. Why it is not obvious ordinarily vis-a-vis the zombie hypothesis is that the hypothesis is not presented in terms of these two mutually dependent perspectives or understandings. Consider Chalmers' presentation of the hypothesis for example:

> ...consider my zombie twin. This creature is molecule-for-molecule identical to me,...but he lacks conscious experience entirely...To fix the ideas, we can imagine right now that I am gazing out the window, experiencing some nice green sensations from seeing the trees outside, having pleasant taste experiences through munching on a chocolate bar,...What is going on in my zombie twin? He is physically identical to me, and we may as well suppose that he is embedded in an identical environment. He will be identical to me *functionally*... He will be *psychologically* identical to me...It is just that none of this functioning will be accompanied by any real conscious experience. There will be no phenomenal feel. There is nothing that it is like to be a zombie (1996, 94).

Here we see that Chalmers equates consciousness with consciousness$_P$ alone, that is, he has in mind only a phenomenological understanding of consciousness. Thus, according to Chalmers, when we attribute consciousness to ourselves or to others we are not thinking in terms of behaviour and physiology – these are at best symptoms of our being conscious. By this measure, we are not *justified* in terms of a person's behaviour and her physiological characteristics to hold that someone is conscious, i.e., conscious$_p$, hence the apparent possibility of a zombie.

But in that case the zombie world Chalmers imagines may be a lot closer to the actual world than he appreciates. If behaviour and physiology are not sufficient evidence for consciousness, then we are, or rather I am, not justified in thinking that others are conscious. I genuinely could not know that others are conscious. I could be living among zombies, nothing rules out this possibility.[45] But even if this were a logical possibility, still, I *know* that others are conscious; and were I to deny that I know this, there would be good reason to doubt that I am fully rational.

Now, Chalmers and other zombists might reply that it is indeed irrational to deny that other people are conscious even though we cannot strictly speaking know that this is the

---

[45] Chalmers rules out this possibility by insisting that consciousness supervenes naturally in the actual world. He explains that this "weaker variety of supervenience arises when two sets of properties are systematically and perfectly *correlated* in the natural world" (1996, 36). He gives as an example how the pressure of a mole of (ideal) gas supervenes naturally on a given temperature and volume, as described by Boyle's law, i.e., pV = KT. This relation between these properties of a gas is not the stronger *logical* supervenience since it is logically possible that K's value (Boyle's constant) could be different such that for a given pressure and volume of a mole of gas the temperature would be different. However, in the case of the consciousness, *understood purely phenomenologically*, we cannot assume such a relation of natural supervenience on a person's physical properties because consciousness so understood is not observable.

case.[46] This, as we have seen, is essentially Jackson's response to the problem of other minds; where he remarks that scepticism about other minds can be dismissed given that the assumption that others are not minded is very implausible (see Jackson 1984, 294). But knowledge is an achievement – at least such propositional knowledge – quite generally we cannot make a knowledge claim without justification. Yet, independent of a naturalistic perspective on consciousness, we cannot justifiably believe that it is true that others are conscious phenomenologically since there can be no evidence for this belief. And again, this fact relates to Nagel's assumption that bats are conscious. He claims that we know that there is something that it is like to be a bat, i.e, that bats are conscious phenomenologically, and yet he claims we cannot know what it is like to be a bat. If someone were similarly to claim that he knows that Peter is tall but he has no idea of Peter's actual height, i.e., he does not know that Peter is not one metre high for example, we would regard his claim as contradictory. For it cannot be the case that someone knows that Peter is tall but does not know that Peter is not one metre high – assuming that Peter is an adult of course. The only way to save Nagel from such a contradiction is to suppose that we know that bats are conscious in virtue of their behaviour and physiology, but we cannot know what it is like to be bat.

We see that how the zombie hypothesis has traditionally been presented assumes we possess only one way of understanding consciousness, namely, as it is understood from

---

[46] The assumption behind this reply is that consciousness can legitimately be thought of in strictly phenomenological terms, and so we must accept as a genuine possibility that we do not know that others are conscious. My point is that we *do* know that others are conscious, and this fact indicates that we do not think of consciousness in strictly phenomenological terms, as evidenced by our not being able to acquire a purely phenomenological concept of consciousness on its own.

the first-person viewpoint. Accordingly, behaviour and physiology have no bearing on whether someone is conscious or not; hence the possibility of zombies cannot be ruled out. The position I have outlined makes the zombie hypothesis incoherent. It is incoherent because a zombie is essentially a person who is justifiably conscious_N, e.g., your zombie twin, and understanding someone as conscious_N is dependent on understanding him as conscious_P at the same time.

The zombie hypothesis is superficially plausible in that we are tempted to think of consciousness only from the first-person viewpoint. That is to say, we are inclined not to suppose that we possess a third-personal understanding of being conscious since consciousness for each of us is primarily equated with its phenomenological qualities. But how we use the term 'conscious' dictates that we possess a third-personal understanding. Only by being justified in attributing consciousness to others do we have a fully shared notion of consciousness. And we can only justifiably attribute consciousness to others in terms of observable evidence, be it direct or indirect, i.e., in terms of behaviour and physiological characteristics. I, like everyone else, believe that others are conscious, but this belief can only qualify as knowledge if there is some way of justifying it as true. If I were to understand consciousness wholly in phenomenological terms, then I would never be justified in this belief since there could be no evidence that others are conscious. But I do know that others are conscious.[47]

[47] Still, the assertion that we *do* know that others are conscious might seem unwarranted by some on the grounds that the possibility that others are not conscious cannot be ruled out, à la scepticism about other minds. Clearly we do imagine that we can conceive of the possibility that others are really not conscious phenomenologically. But this very thought requires grasping the concept of consciousness and this is impossible unless we also think of consciousness in terms of behaviour and physiology. We cannot

4.2.3 The Problem of the Explanatory Gap

I have argued that that the concept of consciousness has two aspects to it, what I have called consciousness$_P$ and consciousness$_N$, which together constitute a single concept because of their mutual dependence. In general, concepts are individuated on the basis of their independence from each other, that is, the statements they are employed in have distinct truth conditions. So the mutual dependence of consciousness$_P$ and consciousness$_N$ implies that when someone asserts 'S is conscious', for example, its truth conditions involve both aspects of consciousness, that is, 'S is conscious' is true if and only if S is conscious$_P$ and conscious$_N$. Here 'S is conscious$_P$ and conscious$_N$' should be distinguished from 'S is conscious$_P$ and S is conscious$_N$' where the latter entails treating consciousness$_P$ and consciousness$_N$ as distinct concepts and the former does not. Formally this distinction

---

even entertain this doubt unless we think of consciousness in this way too. Such doubt consists in the thought 'I know that I am conscious because *this* is what it is to be conscious, but I cannot know that others are conscious in *this* way.' But to think of consciousness as *this* is to think of it in a way that no one else can share. The question 'How can I know that others are conscious like *this*?' has no answer. If you ask yourself this question there is no answer since you cannot in principle confirm or disconfirm the possibility that others are conscious like *this*. If you ask others this question they cannot answer since they cannot in principle know what you mean by '*this*'.

Going back to Wittgenstein's beetle-in-a-box example, the owner of a box can understand beetle as both the contents of the box and as *this*, i.e., the thing she alone sees inside the box. If asked how do we know there is a beetle in someone else's box we can straightforwardly reply that there must be a beetle given that beetle is thought of as what is in the box, *even if the box has nothing in it.* Accordingly, to doubt there is a beetle in the box makes no sense. The question seems to make sense if we think of a beetle not as the thing in the box but strictly as *this*, i.e., as that which only I can see, as a member of the box-owning community. But that is not what 'beetle' means. To use the term 'beetle' requires thinking of it as the contents of a box. The term 'consciousness' is like the term 'beetle' in this respect. And importantly 'beetle' does not simply mean , i.e., refer to, the box. That is because the concept could not have originated unless there had been something in the box.

might seem to be representable by the difference between 'Fx & Gx' and '(F & G)x', but this latter formula is not well-formed. The ill-suitedness of this distinction to description reflects the peculiar or unique nature of the relation between $consciousness_P$ and $consciousness_N$. These aspects of consciousness are best understood as concerning two *perspectives*, namely, how we understand consciousness as a property pertaining to objects, $consciousness_N$, and how we understand it as the property that we, as phenomenal subjects, are manifestations of. And our grasp of the concept depends on having both these perspectives.

However, the conclusion that these two aspects of consciousness cannot be thought of apart from each other will strike many as implausible. After all, this conclusion suggests that $consciousness_P$ and $consciousness_N$ are aspects of, or ways of thinking about, the same property. But, we cannot understand how they could concern the same property. The best way to illustrate why is by considering the problem of the explanatory gap.[48] The explanatory gap describes a crucial epistemological difference between identity statements about physical entities, e.g., 'water is H_2O', and psychophysical identity statements such as 'pain is C-fibre stimulation'. Physical identity statements can help us explain what something is. For example, understanding that water is H_2O explains why water has certain properties such as being an agent in various chemical reactions. By contrast, thinking of pain as C-fibre stimulation does not explain what pain is, that is, it does not allow us to understand why our pain experiences have the phenomenological quality they do. This is obviously a facet of the problem of consciousness since it

[48] See Levine 1983

concerns our inability to understand how consciousness understood phenomenologically is identical with some physical property, as physicalism assumes it is.

Now, to assert that to think of consciousness phenomenologically is also to think of it in physical terms, as is suggested above, would be to assume the explanatory gap is essentially illusory. If to think of pain phenomenologically is also to think of it in physical terms – in ways that make it reducible to C-fibre stimulation for example – then the identity statement 'pain is C-fibre stimulation' should be transparent. That is, once we understand that such-and-such behaviour and physiology, that we think of pain as being, essentially reduces to C-fibre stimulation, we will see that the identity is true. But as the explanatory gap attests, this clearly is not the case. Therefore, consciousness_P and consciousness_N cannot be thought of as concerning the same property.

This objection gains purchase so long as it is possible to assume it is true that the phenomenological quality of an experience is identical with some behavioural or physiological characteristic of the experience. But, to think of consciousness_P and consciousness_N as concerning the same *property* is already to think of them as independent concepts. My claim is that they are aspects of the same concept. This is not to say anything about how consciousness_P and consciousness_N concern the same property.

But, more importantly, I have argued that qualia are unindividuable. Therefore, we cannot in fact identify the quale of some type of experience as any of the physical characteristics of that type of experience. In general an identity statement '$a = b$' is truth-evaluable only if its terms' references are determinable. If one of these terms has an indeterminate reference then the statement can be neither true nor false. And since qualia are unindividuable no term can be employed to refer to any quale-type, e.g., a pain quale.

Therefore, we can never judge a psychophysical identity statement to be true or false, contrary to the assumption motivating the explanatory gap.

But, it is fair to ask why we cannot see straightaway that a term we use to refer to the phenomenological quality of an experience fails to refer – such terms seem to refer. To answer this question it is helpful to revisit Carruthers' example of a recognitional concept. The example concerns a chicken-sexer Carruthers calls 'Mary' who can consistently separate chicks into two distinct types, *A* and *B*, which correspond reliably with the chicks' sexes. It is stated that Mary cannot *explain* how she is able to distinguish *A*-hood from *B*-hood, rather these concepts are purely recognitional for her, i.e., her grasp of them is immediate or intuitive. Qualia, as Carruthers notes, are likewise assumed to be recognitional concepts. To the extent that we are supposed to be able to distinguish a pain quale from an itch quale, for example, we cannot explain how we are able to do so. With respect to *A*-hood Carruthers observes that we would not suppose that *A* refers to a non-physical property despite Mary's being unable to identify it with any physical properties of chicks. Indeed, according to Carruthers it is clear that *A* is the property of being a male chick.

But why should Mary herself identify *A*-hood as maleness or any other physical property of the chick? To her, the one in possession of the concept of *A*-hood, there is no reason to assume that they are the same. Carruthers states that to reason so would be fallacious and that we know that "the property picked out by her recognitional concept is in fact the property of being male" (2001, 57). Certainly, Carruthers is justified in assuming this in that Mary would judge a chick to be *A* by observation, albeit guided by intuition, hence it is absurd not to assume that what she 'subconsciously' detects is a

physical, i.e., observable, property of the chick. But let us focus our concern simply on whether *A*-hood is identical with the specific physical property of maleness. Why must Mary assume that this is the case? Why could she not think instead that *A*-hood and maleness are correlated but that they are not the same property? Mary could think this, but her problem is that no one else would know what she means by *A* in that case. If she were to insist that by *A* she does not mean male chicks but something else, the rest of us would be left in the dark. By assuming *A*-hood is identical with maleness, on the other hand, she is able to use the term '*A*'. '*A*' is then synonymous with 'male chick'. This describes what actually happens with respect to chicken sexers.

But let us now change the situation in one important way and imagine that everyone has an ability to distinguish between *A*-hood and *B*-hood, i.e., these become universal recognitional concepts. By this measure it would seem to be no longer imperative for Mary, or anyone else, to suppose that *A*-hood is identical with maleness since others would 'understand' what she means by *A*-hood without having to identify it as maleness. But what would *everyone* understand by *A* under these circumstances? If they assume it is not some physical property of chicks, i.e., maleness, then it is unclear what the term refers to. At best each of us would be able to say that *A* refers to 'you-know-what', i.e., that thing we each detect some chicks to have. Over all this is a useless concept since there would be nothing to determine that we each mean the same thing by *A*. Accordingly, we could not use the term '*A*' so 'understood' to communicate anything. *We* would not in fact understand anything by the term.

Now, in the case of qualia, understood as universal recognitional concepts, they are not observable. We do not grasp the concept of pain qualia, say, by intuitively detecting

some physical property of pain experiences. Therefore, we seem free of any rational obligation to assume that qualia must be physical properties in the same sense as we would be obligated to think of *A*-hood as a physical property. More importantly, as *universal* recognitional concepts, it would appear that we do not have to identify them with some physical property of experience in order that others may understand what we each mean by the term 'pain' in this sense. But just as in the case of *A*-hood being a universal concept in this sense the term 'pain' could mean nothing to *us*, i.e., it amounts to a useless concept. But we are tempted to think of such terms as being meaningful in the sense that everyone claims to understand what the term means because it has the air of universality – everyone reports that their experiences have some phenomenological quality. It is this air of universality that creates the illusion that qualia terms do refer.

We can again illustrate this point by using Wittgenstein's example of the beetle-in-the-box (1958, 293). We are told that the owner of each box is able to see what is inside it, but that no one else can know its content. Because everyone has a box and the term 'beetle' refers to its content, whatever it may be, *they* can use the term 'beetle' to mean roughly the 'thing-in-the-box'. But the term cannot, and does not, refer to the 'thing' in the box, since no one else but the utterer of the term could know what this thing is. The term 'beetle' can at most mean whatever is in the box given that our concepts are shared and that is all the speakers share across their experiences. The fact that everyone can confidently use the term 'beetle' tempts them each to think that what the term *really* refers to is the thing they see in the box. But the term understood in this way is useless since their experience cannot be shared by anyone else. Likewise, by the term 'pain', for example, we are tempted to think that we each essentially refer to the phenomenological

quality of this experience, since we can all confidently use the term and we each have direct acquaintance with this quality. But, as we have seen, pain can only constitute a concept to the extent that we can all use it, and what we *share* vis-a-vis pain is not its phenomenological quality but the experience's observable characteristics as manifested behaviourally and physiologically. These characteristics play the same role as the box in Wittgenstein's example, that is, they are the means by which we can conceive of such an experience at all. That is not to suggest that the phenomenological quality of an experience is irrelevant to our understanding of an experience; rather, it is to note that pain or any other experience cannot be thought of independently of its observable physical characteristics, even though we are tempted to think otherwise.

Something more needs to be said about the claim that the conceptual inseparability of consciousness_P and consciousness_N does not necessarily imply that they concern the same property. The relation between these two aspects of consciousness should not be thought of as one of identity, understood as a metaphysical relation. One thing can only be judged to be identical with another when they fall under a common concept. For example, we can judge that Mark Twain is identical with Samuel Clemens in virtue of the fact that both singular terms refer to the same *person*. 'Person' is a sortal predicate in this context that allows us to count individuals, i.e., persons, so that then we can judge whether or not two singular terms refer to the same individual according to an identity statement.[49] Thus we cannot, for example, make sense of the identity statement 'Julius Caesar is the number 17' given that we cannot determine a sortal predicate under which both these terms fall. This example concerns what is referred to as the 'Caesar Problem', originally expounded

[49] See P.F. Strawson 1959, 168-175.

by Gottlob Frege. The problem, as Frege saw it, is that we do not have a definition of number to allow us to decide if 'Julius Caesar' is the name of a number or not (Frege, 68). A definition of number would give us a principle by which to determine the sortal predicate *number* such that we can exclude Julius Caesar, the conqueror of Gaul, from falling under it. Frege thought our inability to refute such identity statements as the one above is a problem because he assumed that singular terms like 'Caesar' and '17' pick out individuals however we conceive the world. Or as Paul Benacerraf explains it: "To speak from Frege's standpoint, there is a world of objects – that is, the designata or referents of names, descriptions, and so forth – in which the identity relation has free reign" (1982, 286). Thus, Frege held that such identity statements must be either true or false, i.e., truth-evaluable, and insofar as we cannot evaluate them this is a result of our conceptual shortcomings.

This line of reasoning, notably, echoes Nagel's complaint about our inability to explain how phenomenological qualities, i.e., the what-it-is-like of experiences, are identical with physical properties. Nagel tells us that we require a *conceptual* bridge between the first- and third-person viewpoints to resolve this the problem of consciousness (Nagel 1974, 449). But, as I stated earlier (section 4.1), we can identify properties only if they fall under a concept; that is to say, we cannot think of properties outside of some theoretical framework. In general, therefore, in order to determine if some property *F* is identical with or distinct from some property *G* we need to have grasped the sortal predicate *property*. But this we can do, stating that a property $=_{df}$ a characteristic or attribute of an entity. Accordingly, it seems to make sense to ask if qualia, as properties so defined, are identical with some physical properties. But

importantly, such judgments are only possible when these properties themselves fall under a concept such that they are individuable. Ordinarily this is not a problem. For example, blue is a property in virtue of its falling under the concept of colour. Qualia, however, are not determined to be properties in virtue of falling under some theory, that is, as posits of some theory. Rather, qualia are properties *simpliciter* with which we are acquainted without their being posited because of our special relationship to them. As has been noted by many commentators, e.g., Dretske and Rowlands, what is peculiar about qualia is that they are the properties with which we apprehend things in the world. Again, qualia are special properties, therefore, because they are the properties that in some sense are constitutive of us rather than properties we apprehend.

4.3 Making Sense of Qualia Talk

There remains one worry that needs to be addressed. If it is assumed that qualia are unindividuable such that the term cannot be used in any verifiable statements, then we are left to conclude that such statements are meaningless. But this seems plainly wrong. For example, the statement 'I am now having a red quale looking at this ripe tomato' seems both meaningful and truth-evaluable. I agree. There is a sense in which such statements are verifiable. Someone could observe, for example, that the tomato I am looking at is indeed red and this would verify my claim, or if this witness were to observe that it is in fact green, say, the claim is disconfirmed. The claim is confirmable, therefore, insofar as it is *broadly* interpreted as saying something along the following lines: 'I am now looking at this tomato, and there is something that it is like to have this experience'. This is not to suppose that what this something is is identifiable.

But, of course, a likely response to this claim is that my having a red quale is not logically entailed by the object I am looking at being red. It is conceivable that even if the tomato is in fact red I am not having a red quale. This is simply another way of expressing something like the inverted spectrum hypothesis. Therefore, these kinds of statements about qualia, interpreted in this *narrow* sense, are not verifiable on these grounds. But our concern, recall, is to determine whether holding that qualia are unindividuable entails holding that qualia do not exist. As such, to deny this entailment only requires that the statement can be plausibly interpreted in the original broad sense, whereby it is verifiable. Because, according to this interpretation, it is supposed that there is something that it is like for me to see red. And it seems clear that this statement *can* be plausibly interpreted in this broad sense.

4.4 Summary

The problem was how to maintain that qualia are unindividuable, and hence unnaturalisable, given that I have argued that they are properties of consciousness and that consciousness is naturalisable. These appear to be contradictory claims. My reply has been that we cannot think of consciousness understood in terms of qualia independently of our understanding of it in terms of behaviour and physiology, i.e., our naturalistic understanding of consciousness. Accordingly, qualia are aspects of the same concept of consciousness that we understand in naturalistic terms. Nevertheless, this is to stop short of holding that qualia are identical with certain naturalistic properties of consciousness. That is because to assert such identities requires there to be a theory spanning both these

aspects of consciousness, and no such theory can exist given that qualia are beyond, or more precisely at the limit of, our theories as indicated by their unindividuability.

In addition, in the next chapter I argue that we can understand how we are acquainted with qualia while being unable to subsume them under our scientific theories so long as we think of them as the properties in virtue of which we apprehend things, rather than properties we apprehend. In order to subsume some properties under our theories, after all, there must be some means of confirming statements concerning them as posits of some theory, which requires them to be either directly or indirectly observable. And as properties by which we observe things qualia cannot themselves be observed.

Chapter 5

Demystifying Qualia

The purpose of this concluding chapter is to consider two difficulties that present themselves for my view of the problem of consciousness and to summarise how I think qualia can be accommodated from a naturalistic perspective. First, I outline two possible objections to my position; namely, its leading to the endorsement of property dualism and its implying a troubling form of mysterianism. These objections are replied to in sections 5.2 and 5.3 respectively. Then in section 5.3.1 I follow up on an issue related to the second objection; this concerns the worry that because qualia are unnaturalisable neuroscience can have nothing to say about consciousness. In reply, I argue that because our understanding of consciousness in terms of qualia, i.e., phenomenologically, is conceptually inseparable from our understanding of it in physiological and behavioural terms neuroscience is relevant to the study of consciousness. Moreover, I argue that the unnaturalisability of qualia is unproblematic so long as we think of qualia as what I call *epistemically originating properties*. And in section 5.4 I outline more precisely what this conception of qualia amounts to.

In my discussion of the relevance of neuroscience to the study of consciousness the idea of a zombie is reintroduced. Now, I had concluded that the idea of a zombie is conceptually incoherent, that is, we cannot think of a person physically and behaviourally indistinguishable from you or me who is non-conscious phenomenologically. However, the notion of zombiehood we shall focus on relates to Dennett's example of zombie Martians, i.e., alien automata, studying human consciousness. This creature is a zombie in the sense of being phenomenally non-conscious but nonetheless able to communicate with us, while not being physically indistinguishable from us of course. I argue that there would have to be a behavioural difference between us and these Martian automata, namely, that in this Martian speech community statements about what it is like to have some experience would never be expressed. This linguistic difference between us and them would imply that automata, i.e., non-conscious creatures, could not originate the concept of phenomenal consciousness.

However, in section 5.4 I argue that the nature of zombiehood in this sense would be such that they could not conceive of much at all. Here, it is worth noting that the possibility of zombies in the fullest sense, i.e., as non-conscious creatures physically and behaviourally indistinguishable from us, is rarely considered beyond their intuitive plausibility. So in this section I take on the idea of zombies on the assumption that there is nothing that it is like to be such a being. This crucial fact, I suggest, implies that they would lack a self, i.e., subjecthood, such that it is very difficult to think of them as epistemic *agents*. The aim of the relatively detailed discussion about zombies that follows is to elucidate the notion of qualia as epistemically originating properties. I end the

chapter with a summary of my overall position, suggesting that the problem of consciousness can be dissolved.

5.1 Two Objections

I have argued that qualia are real, i.e., they do exist. However, conjoining this conclusion with the claim that qualia are unnaturalisable would suggest that qualia, and *a fortiori* consciousness, are, to paraphrase Gilbert Ryle, mysteriously occult properties.[50] Moreover, if qualia cannot in principle be captured or subsumed under our scientific theories then there is no reason to assume they fall under the laws of nature, i.e., the laws of physics. Therefore, while consciousness is a property of certain kinds of physical things such as human beings, it is not itself physical. Accordingly, my view seems to collapse into property dualism.

If, on the other hand, qualia were naturalisable, then there would be the possibility at least that by observing a creature's brain in operation we could determine whether it is phenomenally conscious. It seems, therefore, I must deny this possibility on the grounds that we cannot ever relate the physical properties of the brain to the properties of conscious experiences, i.e., qualia, since qualia are unnaturalisable. Thus, the practice of observing and investigating the physical processes of the brain, i.e., the practice of neurophysiology or neuroscience more generally, has nothing to do with consciousness understood in phenomenological terms. Jaegwon Kim acknowledges a similar kind of

---

[50] The worry is that I am committing what Ryle famously called a category mistake, namely, I am treating qualia as entities that exist in the same way as ordinary physical properties. Ryle points out that thinking of the mind, as he puts it, as *being* like the body is like thinking of the average taxpayer as someone who one could meet on the street. This way of thinking of the mind makes it seem like a wholly mysterious entity (see Ryle 1949, 19).

difficulty for his own view, which he articulates in terms of the problem of trying to engineer a machine that feels pain.[51] Kim argues that pain experiences are functionalisable, and hence subsumable under our theories in terms of physiology and behaviour, but that their phenomenological quality is not functionalisable. He states: "We can...easily design into a machine a device that will serve as a causal intermediary between the physical input and the behavioral output" (2005, 168). However, he adds that we cannot begin to understand the connection between the causal function of such a machine and the realisation of pain qualia when the machine is activated; consequently he concludes that "[t]he machine would try to flee when its skin is punctured even if we had, wittingly or unwittingly, designed itch or tickle into the box [i.e., that part of the machine 'designed' to realise pain qualia]" (*ibid*). The unnaturalisability of qualia in the same way would appear to rule out the possibility of our ever knowing how and when qualia are realised physically.

These worries about my view point to two distinct difficulties. First, there is the danger that the view amounts to property dualism. I seem to be forced to conclude, or at least I cannot defend against the charge, that qualia are non-physical properties; and with this come the usual problems, e.g., epiphenomenalism. Second, my view appears to be strongly mysterianist in tenor. Most starkly expressed the view seems perhaps fatally sceptical – I am led effectively to assert that it is impossible in principle to know if others

[51] For Kim qualia are subsumable under our theories in terms of their relational characteristics as manifested physiologically and behaviourally. But he distinguishes these characteristics from the felt qualities of conscious experiences which he argues are not subsumable. Kim distinguishes between the physical and phenomenological aspects of consciousness because he equates qualia with phenomenal *states* (see Kim 2005, 10-11).

are conscious understood strictly in phenomenological terms. I shall next reply to these difficulties in turn.

5.2 Why Qualia Are not Non-Physical

The immediate worry about adopting a naturalistic attitude is that it seems antithetical to physicalism. Physicalism, after all, is a claim about the world as such, i.e., reality. It is a metaphysical doctrine based on the thesis that everything that exists is physical. One might say that physicalism is not a hypothesis about how the world is in some respect, but rather it is a hypothesis about how it is in the most general sense. Thus, it is hard to see how it can be viewed as a scientific hypothesis – to the extent that it does not concern any specific features in the world.

But, any scientific hypothesis is held true so long as it is not contrary to our observations in conjunction with our theories. By this measure physicalism *can* be straightforwardly thought of as a scientific hypothesis. We can assume that everything is determined by the physical facts so long as there is no evidence against this assumption. What accordingly might falsify physicalism? One way Quine imagines is if there were convincing evidence of extrasensory perception. Imagine, for example, some experiment were to indicate that thoughts are transferred reliably, i.e., at a far greater success rate than can be explained by chance, between two physically isolated persons. If this were to happen then we would perhaps have good reason to think that not all causation is physical. Accordingly, these 'thoughts' could in effect be thought of as non-physical in that their interaction would not be governed by the laws of physics. That said, Quine notes that in such circumstances "it would still not devolve upon psychologists to

supplement physics with an irreducibly psychological annex. It would devolve upon the physicist to go back to the drawing board and have another try at full coverage, which is his business" (1998, 430). Quine's point is that physics is universal in scope, that is, its aim or business is to conceive of every phenomenon in its own right. Therefore, if some phenomenon, such as extrasensory perception, were to fall outside of our present physical theories this would not in itself show that it is non-physical; instead this aberrant phenomenon likely would tell us that our physical theories need revising, given that physics covers everything. Only if we were unable to incorporate this aberrant phenomenon in our physics could we conclude *as a last resort* that it is non-physical, and consequently that physicalism is false.

Moreover, Quine takes physicalism to rest on naturalism; more precisely, he writes: "I do embrace physicalism as a scientific position, but I could be dissuaded of it on future scientific grounds without being dissuaded of naturalism" (2004, 281). It is in this way that physicalism can be thought of as a scientific hypothesis. What makes the hypothesis 'metaphysical' is its generality. Again, it is a claim about everything, i.e., all phenomena, rather than about certain aspects of the natural world as conceived according to our particular interests, as in the case of higher sciences such as biology.

This naturalistic conception of physicalism can be summed up as follows: everything is physical, but this is a scientific hypothesis that could be falsified by the right kind of evidence. This evidence would have to be inconsistent with our scientific theories and require revisions to them in order to overcome this inconsistency. Moreover, the least radical way of revising our theories would have to be to abandon the physicalist hypothesis. If such evidence were to present itself, then we would be led to accept

dualism. This dualism would amount to the claim that some phenomena are not governed by the laws of physics as we understand them. Let us call the resultant dualism *naturalistic dualism* (not to be confused with Chalmers' position which he gives the same name). Certainly, such dualism would be radical inasmuch as, initially at least, there would be no theory enabling us to explain the phenomena concerned. There would suddenly be a conspicuous gap in our scientific theories. However, naturalistic dualism would not threaten our scientific worldview in that it would itself be a scientific hypothesis. We would still be able to suppose that the dualist hypothesis could *in principle* be incorporated into our theories. This possibility would still be left open, however difficult this might seem to be to do.

That said, it would be very difficult to imagine such a dualism since it would be contrary to the causal closure of the physical. We could not assume mental phenomena are epiphenomenal because the evidence against physicalism that we imagined was the transference of thoughts non-physically from person *A* to person *B*, which would imply that thoughts are non-physical but detectable behaviourally. Epiphenomenalism, of course, would not allow us to suppose that *A*'s thoughts effectively caused *B*'s behaviour, as would be the case. And giving up the principle of causal closure of the physical is barely conceivable. In terms of Quine's maxim of minimum mutilation rescinding physicalism is very much a last resort, close to calamitous vis-à-vis maintaining the overall coherence of our theories.[52]

Importantly, naturalistic dualism must be distinguished from what I call *non-naturalistic* dualism. According to this dualism there are phenomena, e.g., qualia, that

[52] See Quine 1992, 14-15.

cannot in principle be incorporated into our scientific theories. This is the brand of dualism advanced by the early Jackson and Chalmers. Here, Chalmers is a non-naturalistic dualist because, as noted earlier, the reforms to science he envisages, that would allow us to incorporate consciousness as he understands it, are intolerable (see 3.2). These reforms would involve abandoning the requirement that any natural scientific hypothesis be empirically testable, either directly or indirectly. Such a reformed science would not be science at all. And the early Jackson's non-naturalistic dualism is betrayed, for example, in the introductory remarks to his 'Epiphenomenal Qualia', where he writes:

> I think there are certain features of bodily sensations especially [i.e., qualia],...which no amount of purely physical information includes. Tell me everything physical there is to tell about what is going on in a living brain,...and be I as clever as can be in fitting it all together, you can't have told me about the hurtfulness of pain, the itchiness of itches..."(1982, 127).

In other words, he claims that it is impossible to relate qualia to physical facts, that is, facts that are describable in physical terms and hence incorporable into our theories quite generally. He takes qualia to be non-physical and epiphenomenal in this sense. The promise of incorporating qualia into the rest of our theories is ruled out – it can *never* happen, according to Jackson.

The worry is that my position entails this more problematic non-naturalistic dualism. Qualia, I have maintained, are in principle unnaturalisable. But, if qualia cannot in

principle be subsumed under our scientific theories, then there is no reason to assume they fall under the laws of nature and hence under the laws of physics. Therefore, my position not only seems compatible with this dualism, it even suggests it. But, this property dualism is only entailed, I argue, if qualia are thought of as individuable. Let me explain.

Implicit in this objection is the following conditional: for every phenomenon, if it is physical then it is naturalisable. And since I claim that qualia are unnaturalisable, it follows *modus tollens* that qualia are non-physical. Now, this conditional is arguable. The conditional is inapplicable to qualia. The conditional would only apply if qualia were individuable; and because they are *un*individuable, there is no sense in which we can think of these properties as falling or not falling under the laws of physics. Another way of putting this is to say that qualia are not candidates for naturalisation; they are not posits in our theories, such as properties like being red, being neuronal, having a length, etc.

The claim above might however elicit the response that if qualia are not entities in any sense, i.e., identifiable and thereby quantifiable as individual properties, then we cannot understand them as real, as existing in the world. It seems qualia have been eliminated. But, what marks out qualia as properties is their self-evident existence, that is, the lack of need to justify – or indeed the impossibility of justifying – our believing they exist. One might put it this way: qualia exist in virtue of there being a (phenomenal) world for each of us. In other words, none of us can be said to know that qualia exist, or more generally to know that we are conscious phenomenologically, in the sense of being able to justify the belief as true. That is because qualia are the properties by which we apprehend things in the world, i.e., the things we ultimately posit by our theories. Their existence is a

precondition for the possibility of science, i.e., knowledge about the natural world. Again, this is to understand qualia as epistemically originating properties.[53] At best we can say that my belief that qualia exist is evidenced by the very act of believing, that is, by my being a subject rather than there being nothing for me, i.e., there being no me. More will be said about this shortly. It is in this sense that the existence of qualia is self-evident and we do not need a theory by which to posit them.[54] Only entities positable in our theories, very broadly construed, can be physical *or non-physical*. Qualia are not such entities; hence there is no sense in which we can think of them as being non-physical, as the property dualist thinks.

5.3 Dissolving the Mystery

Qualia, then, are not candidates for naturalisation. However, this conclusion still seems to leave us with a sense of mystery vis-a-vis qualia – they are properties with which we are each intimately acquainted but about which we can say nothing. They appear to be mysterious because they are unnaturalisable. It would seem, therefore, that my position is a form of mysterianism. Interestingly, Seager likewise suggests that consciousness – understood here in purely phenomenological terms – must be presupposed in any

---

[53] The idea of a property being epistemically originating does not concern thinking of these properties as given in the sense of being indubitable, i.e., as epistemically foundational, such that all our knowledge is ultimately *justified* by our realising them. Rather, to reiterate, the idea is that the realisation of these properties is the precondition for knowledge in that without realising them we cannot apprehend the world in contradistinction to ourselves as knowers. More is said about this later in the chapter.

[54] The term 'theory' is used in a very broad sense. Our common sense knowledge concerning such things as what is a table should be thought of as continuous with science. Therefore, this common sense knowledge can be construed as theoretical in nature – we have the concept of table, namely, as a piece of furniture that fulfils a certain set of

naturalisation of a phenomenon and so cannot itself be naturalised. This amounts to what Seager calls a kind of *methodological mysterianism*. Consciousness is mysterious, i.e., unnaturalisable, by this measure because the conditions for naturalisation, as he defines it in terms of the three rules noted earlier (see section 1.1), preclude our being able to naturalise the mind or any aspects of it, including consciousness.[55] Such mysterianism, he suggests, best fits with the kind of constructive empiricism developed by Bas van Fraassen. Very roughly van Fraassen's view is antirealist or instrumentalist. He argues that we cannot know if any of our scientific theories describe reality, rather at best we can only commit to these theories to the extent that they are empirically adequate, i.e., they are predictively successful.

Constructive empiricism is indeed compatible with mysterianism since it assumes that there is some ultimate reality and we cannot know if our scientific theories describe it. But by contrast the naturalism I have urged is not antirealist. In this regard I again follow Quine. Reality is what our best scientific theories describe. There is no transcendent viewpoint from which to grasp some ultimate reality that our theories strive to describe. What counts as real is what we posit in our theories. Quine explains:

> To call a posit a posit is not to patronize it. A posit can be unavoidable except at the cost of other no less artificial expedients. Everything to which we concede existence is a posit from a standpoint of a description of the theory-building process, and simultaneously real from the standpoint of the theory that is being built. Nor let us look down on the standpoint of the theory as make-believe; for we can never do

---

functions etc. Theories about such everyday items are of course less rigorous than scientific ones, hence the distinction between common sense knowledge and science.

> better than occupy the standpoint of some theory or other, the best we can muster at the time (1960, 22).

Naturalism, therefore, leaves no room for mysterianism. A helpful way to illustrate this naturalistic attitude is to follow Quine in using Otto Neurath's metaphor of a boat at sea.[56] Science, as a boat on the high seas, is necessarily built piece-by-piece so that it can remain afloat. We cannot take our science into a drydock so to speak and build it anew from top to bottom all at once. In other words, there is no external or transcendent viewpoint from which we can evaluate our scientific claims *in toto*. Again, there is no higher authority than our senses to appeal to with respect to judging our claims as true or false; and since this is the authority that grounds our scientific claims and the theories that support them, philosophical claims must be consistent with them, i.e., with science. Thus, as Quine remarks, the philosopher's task differs in detail from others, e.g., the scientist's, but still it differs "in no drastic way as those suppose who imagine for the philosopher a vantage point outside the conceptual scheme that he takes in charge. There is no such cosmic exile" (1960, 275). That is to say, there is no essential difference between the perspective of the philosopher and of the scientist.

To put this another way, the project of science in this inclusive sense necessarily involves constructing theories piece-by-piece rather than all at once, since to construct our theories all at once, or anew, would require our being in the position to evaluate our

---

[55] See Seager 2000, 95-129.

[56] In his book *Word and Object* Quine quotes Neurath as follows: "Wie Schiffer sind wir, die auf offener See umbauen müssen, ohne es jemals in einem Dock zerlegen und aus besten Bestandteilen neu errichten zu können" [We are like sailors who must renovate their boat on the open sea, without ever being able to dismantle it in drydock and build it anew from the best components] (1960, vii).

theories within a comprehensive theoretical framework effectively based on a second-order theory that is independent of science. This comprehensive second-order theory would therefore have to presuppose an external vantage point, a way of taking the boat out of the water into drydock. But there is no 'more solid' ground because there is no higher authority by which to appeal to the truth of any theory than that appealed to by science, namely, our senses. Therefore, any such comprehensive theory cannot be independent of science, i.e., cannot be of a higher order. One might say that the test of some theory is based on its ability to stay afloat, that is, on its ultimate agreement with our observations.

Qualia, then, are not unnaturalisable such that they ought to be naturalisable but we are unable to naturalise them. Their unnaturalisabilty does not leave us feeling vexed. They are unnaturalisable in the sense that they are not candidates for naturalisation. Therefore, there is no need for us to feel something has been left out, i.e., left unexplained. Certainly, in the same way we note that our eye cannot gaze upon itself and therefore wonder what our eye looks like we can wonder what qualia are in relation to other phenomena, but such curiosity is superficial. Upon reflection we realise that there can be no way in which the eye gazes upon itself, and likewise we ought to realise that there is no sense in which qualia can be understood in terms of their relation to other phenomena.

Doubtless many will demur at the idea that what we count as real is determined by our best theories; this sounds like idealism. What is real, many argue, is not necessarily graspable by us. We saw this kind of response from Carruthers vis-a-vis his property realism (see section 4.1). Nagel expresses a similiar response with the following remark:

"Realism is *most compelling* when we are forced to recognize the existence of something which we cannot describe or know fully, because it lies beyond the reach of language, proof, evidence, or empirical understanding" (1986, 108, my italics). Of course this belief in evidence transcendent truths is premissed on the assumption that there is a transcendent viewpoint, i.e., an epistemic position from which *everything* can be known in principle by some infinite mind.

It is this compelling form of realism that Hilary Putnam attacks with his famous brain-in-a-vat thought experiment.[57] Putnam imagines someone worrying that we may all in fact be bodiless envatted brains wired up to a mad scientist's vast computer, assuming even that others are not merely programmed into your world so to speak. How are we to undermine this possibility? Putnam's way of undermining it is to argue that the sentence 'I am a brain in a vat' is self-contradictory. An envatted brain cannot refer to brains such as itself; at most by 'brain' it can refer to a certain patterns of signals from the computer, and so the sentence 'I am a brain in a vat' sincerely uttered by an envatted brain is false. On the other hand if someone who utters this sentence is not an envatted brain then the sentence is false as well.

But if, as Nagel suggests, reality is beyond language one might reply that whether each of us really is an envatted brain is independent of our ability to know this or express it. So if in reality, in this robust compelling sense, I am an envatted brain then my inability to express this truth makes no difference. But this inability does make a difference. As Putnam points out, "if we *are* brains in a Vat, we cannot *think* that we are" (1981, 50). In other words, the brain-in-a-vat hypothesis is logically incoherent for us at least. Reflecting on Putnam's observations, we can understand how the hypothesis is

logically incoherent because we are never in the position to confirm or disconfirm it, since this would require per impossible our being able to experience our experiencing the world. Therefore the hypothesis is empty.

In general, because there is no transcendent viewpoint from which to judge whether our theories about how the world is, i.e., our scientific theories, are true or false – no drydock in which we can build our theories anew – we must settle with our theories themselves. There are tables, chairs, *brains*, atoms and so on in virtue of our theories requiring them. Talk of such things existing independently of our theories makes no sense given that they are identified only within them. That said, qualia, despite being unnaturalisable, can be thought of as real because as epistemically originating properties they are the properties by which we are able to apprehend the world, and thereby form theories in the first place. None of our scientific theories would be possible without our having a viewpoint as constituted by qualia.[58]

The original concern was that qualia are unnaturalisable and consequently they are wholly mysterious entities. There is no way in principle that we can ever explain what qualia are, and hence there is no way we can explain consciousness. Thus, to borrow Kim's example, if we were required to *engineer* a device, i.e., design it from scratch, that realised qualia we would have to admit defeat despite qualia being physically realised in some way. In reply, I have argued that from a naturalistic perspective insofar as consciousness is explainable there is nothing more to *know* about it than what our

---

[57] See Putnam 1981, 1-21.

[58] Ted Honderich's notion of consciousness as the existence of *a* world is a useful way of understanding the basis for this thought. Honderich explains: "The difference between me now and a chair in this room ...is that for me a world exists, and for a chair a world does not exist" (2004, 130). In this respect reality might be thought of as our world.

scientific theories tell us in terms of physiology and behaviour. In this regard qualia are not properties that are in principle open to explanation – we should not expect an account of them. But, by this measure one may wonder what the relationship between neuroscience, as the study of the human brain and its cognitive processes, and phenomenology, as the study of consciousness from the first-person viewpoint, is. Are we to suppose that phenomenology is a pseudoscience? That is the question we shall take up next.

5.3.1 Neuroscience and Consciousness

I have said that consciousness is naturalisable to the extent that it is understood as a physical property, that is, as a complex property we attribute to physical bodies in virtue of various physiological and behavioural manifestations. Moreover, I have argued that consciousness is understood *both* in physical and phenomenological terms, and that we cannot think of consciousness without thinking of it in both these ways. These two ways of thinking of consciousness are inseparable – we cannot acquire the concept of consciousness by thinking of it in only one of these two ways.

But again, this is not to claim that qualia, as the phenomenological qualities of conscious experiences, are identical with some physical properties of the brain, say. We cannot claim this because qualia are unindividuable. Identity claims are only possible *within* a theoretical framework that provides us with a sortal concept, and qualia cannot fall under any such theories. Accordingly, we do not *observe* that qualia are correlated with brain properties, such as the firing of neurons, in the way that it was first observed,

---

While the details of Honderich's theory of consciousness are perhaps problematic, this basic insight he offers about consciousness is still valuable.

for example, that the Morning star is correlated with Venus, and thus subsequently shown to be identical with the planet. That would require our being able to pick out our experiences in terms of qualia as such, i.e., independently of the physical manifestations of our experiences.

But despite qualia being unidentifiable with their neural correlates we can still assume consciousness *as we conceive of it* concerns both qualia and brain properties. Again, our concept of consciousness only exists because we can think of our experiences in terms of their behavioural and physiological manifestations *as well.* And this conceptual dependence of consciousness understood phenomenologically on physiological and behavioural properties, and therefore ultimately on brain properties in particular, gives us license to suppose that the study of brain processes is relevant to our phenomenological understanding of consciousness, although how neurophysiology, or neuroscience more generally, informs our understanding of consciousness phenomenologically has not been made clear. Neuroscience, after all, cannot tell us *what* qualia are in the sense of reducing them to such-and-such brain states either as types or even as tokens.

Below, I aim to show that neuroscience does concern our phenomenological understanding of consciousness. To do this I shall consider Dennett's claim that a race of phenomenal zombies could gain a complete or completable understanding of consciousness in terms of what he calls *heterophenomenology* in conjunction with neuroscience. In reply, I argue that what such aliens would come to understand neuroscience to be about would differ from our own understanding. Unlike us they could only understand consciousness in naturalistic terms, i.e., they could only develop an

understanding of consciousness$_N$, and would do so only in virtue of our already possessing the concept of consciousness fully, i.e., understood both as consciousness$_N$ and consciousness$_P$. This fact shows that what we take neuroscience to be about is informed by our being phenomenally conscious creatures.

The worry is that we can study the brain as much as we like, but this will never enable us to understand the essentially subjective aspect of our conscious experiences. Even after the cognitive functions of the brain have been mapped out in detail the neuroscientist will still be at a loss to explain why the brain determines our being phenomenally conscious in this sense. Unless this can be explained, in other words, one might think that neuroscience has *nothing* to say about consciousness as we understand it in phenomenological terms. This harks back to the worry expressed by Nagel that because phenomenological facts concerning our experiences are essentially connected to the first-person point of view there seems to be no possibility of understanding them in terms of properties apprehensible from the third-person point of view (see section 2.1). However, I think we can overcome this concern; that is to say, the worry is misplaced.

Consider, first of all, Dennett's way of allaying this worry. In general, neuroscience needs to take into account consciousness understood phenomenologically. A full understanding of consciousness should include an account of our first-person reports of conscious experiences. We need to make sense of such first-person reports, e.g., 'I have a *sharp* pain in my arm' in neuroscientific terms. But, a neuroscientific account of consciousness in terms of behaviour and physiology would seem to bear no relations to our understanding consciousness from the first-person viewpoint. Phenomenology is usually or traditionally thought of as the study of consciousness *strictly* from the first-

person viewpoint. In this sense it is a theory of appearances, i.e., it aims to explain appearances or seemings as objects in their own right, analysing them purely in terms of their relations to one another. It takes seriously the idea that we can introspect our phenomenal states. This approach to the study of consciousness is perhaps epitomised by Husserl's phenomenology. And as Barry Smith and David Woodruff Smith explain,

> Husserlian phenomenology seeks the description and structural analysis of consciousness, as opposed to an account of its causal origin in brain activity or elsewhere. Consciousness is to be studied precisely as it is experienced, and accordingly the *objects* of consciousness, too, need to be characterized precisely as they are given in experience, with no metaphysical reinterpretations (inspired by reductive or other motives) (1995, 9).

Phenomenology so construed does indeed seem in principle unrelatable to neuroscience, since neuroscience concerns such 'brain activity'. Here we are again reminded of Nagel's assertion that understanding what it is like to be a particular kind of creature requires taking up the viewpoint of that creature, and so if we abandon that viewpoint, i.e., take up a third-person viewpoint, then our understanding of it is lost (1974, 444-445).

However, Dennett denies that phenomenology can only be pursued from a first-person viewpoint. He argues that people's first-person reports are essentially a form of behaviour and as such they are analysable from the third-person viewpoint. People's sincere reports about how things seem to them offer a publicly accessible means of studying consciousness in phenomenological terms. Therefore, phenomenology can be

construed as a fully objective science. This approach to phenomenology Dennett calls 'heterophenomenology', which is contrasted to Husserl's *auto*phenomenology, i.e., the study of consciousness by some sort of self-analysis. The immediate objection to Dennett's idea is that what phenomenology studies is not people's reports *per se* but the things their reports are about, i.e., appearances. The pertinent phenomena are the appearances themselves and not the sentences about them.

Dennett's reply to this objection is twofold. First, he argues that Husserlian phenomenology is flawed. It is based on a person's reports about appearances or phenomena that are assumed to be apprehensible only by that person, i.e., essentially private. Consequently such reports are not objectively confirmable. Moreover, it seems that naturally people tend to believe whatever agrees with their already acquired stock of beliefs – a phenomenon known as confirmation bias. Accordingly, this *auto*phenomenological approach to the study of consciousness does not meet the standards of scientific scrutiny. Second, it is scientifically legitimate to treat such reports *as if* they are true and then determine whether they agree with the empirical evidence. For example, this would be the approach taken by an anthropologist studying a newly discovered tribe. She would treat their myths as true, i.e., not immediately doubt the veracity of their stories and dismiss them, in order to reach an understanding about their *notional* world, that is, their self-consistent model of the world.[59] In other words, judgment is suspended vis-a-vis these myths – this amounts to a form of *epoché*. Likewise, it is useful to think of people's first-person reports about how things seem to them as descriptions of their phenomenological notional worlds. We can then evaluate

[59] See Dennett 1991, 82-83.

such phenomenological notional worlds in terms of how well they agree with our observations.[60]

This is to adopt a strategy based on what Dennett calls an *intentional stance*. The strategy is to attribute a belief to someone, or some intentional system more generally, insofar as it enables us to make sense of or rationalise his actions. So, for example, if someone were to take to wearing a bicorne (a two-cornered hat), speak in French, and refer to his wife as 'Josephine', however improbable, one could rationalise his odd behaviour by attributing to him the belief that he is Napoleon Bonaparte. To the extent that this belief fits with all the evidence one can hold that his believing this is true. It is taken to be true that the man believes this irrespective of whether he *really* believes it in the sense that perhaps he might only be pretending to be Napoleon and never admit to it – so long, that is, as his belief that he is not Napoleon is never made apparent by his behaviour or physiology, i.e., by any empirical evidence. The intentional stance, then, is based on the attribution of beliefs to someone *according to how he ought to behave* under the relevant circumstances, e.g., when asked if he believes he is Napoleon he answers affirmatively; in other words, the attribution is true to the extent that the belief

---

[60] Implicit in Dennett's argument is a Quinean picture of interpretability. That is, he thinks of the heterophenomenologist as essentially being in a position analogous to the field linguist imagined by Quine (Quine 1960). This linguist is challenged to interpret the *utterances* of some native whose language is entirely unknown. So, for example, when the native utters 'Gavagai' the linguist can *only* interpret the native in terms of his observations of the scene being reported on, i.e., a scene with a rabbit in Quine's example. We cannot suppose that in addition the native means something that is closed to observation in any way. As Quine puts it: "All the objective data he [the field linguist] has to go on are the forces he sees impinging on the native's surfaces, and the observable behavior, vocal and otherwise, of the native" (ibid, 28). Likewise, Dennett sees the heterophenomenologist as interpreting people's first-person reports on how things seem to them solely on the basis of observable evidence, i.e., real patterns of behaviour. There is no matter of fact about how to interpret these reports.

successfully predicts his behaviour. What matters is the strategy is very successful and useful in itself. This strategy, Dennett argues, is useful vis-a-vis the study of consciousness.

Thus Dennett's suggestion is to hold as true a subject's first-person reports and beliefs about his phenomenal states and to continue to do so as long as these reports fit with all the observable data. Thus, the reports themselves are taken to be behavioural data. By this measure, the question of whether these phenomenal states *really* occur in the subject, so to speak, is irrelevant. Indeed, Dennett admits that a so-called phenomenal zombie could effectively pass the test in that insofar as it reported having such-and-such phenomenal states and these reports concurred with what we observe in terms of both its physiology and its behaviour otherwise, we would accept its reports about how things seem to it as true, despite its 'experiences' having no qualitative character at all. Dennett takes this as evidence that the idea of a zombie is incoherent – a zombie is in principle indistinguishable from a phenomenally conscious human being.

Our worry was that neuroscience tells us nothing about what it is like to experience things. Another way of viewing this difficulty is to note that neuroscience and phenomenology, while both being ways of understanding consciousness, are unrelatable – neuroscience takes as its data that which we can observe about the brain and cognition while phenomenology takes as its data those appearances we each introspect, and which consequently are inapprehensible from the third-person viewpoint. Dennett's reply to this worry is to suggest that phenomenology can legitimately be construed as the analysis of our first-person *reports* about our phenomenal states – a construal he calls 'heterophenomenology' – and these reports themselves are open to observation, i.e., they

are apprehensible from the third-person viewpoint. Thus, the barrier between these two ways of studying consciousness is removed. Hence we *can* relate our neuroscientific theories to our talk about how things seem to each of us.

Dennett's suggestion is right to the extent that heterophenomenology legitimately allows us to relate consciousness understood phenomenologically to brain processes, ultimately. Indeed, as Dennett notes, it is an approach that has already been adopted by neuroscientists. The trouble with this approach is that it leaves no room for qualia. A virtue of heterophenomenology, according to Dennett, is its metaphysical neutrality, i.e., the fact that it lets us study consciousness understood phenomenologically irrespective of whether there really are qualia or not. This is certainly a virtue so long as one is not concerned to account for qualia – something that Dennett is happy with since he denies that qualia exist.

Nevertheless, Dennett insists that he is not denying that this subjective aspect of experience exists. He is not suggesting we are all 'zombies'. But he is suggesting that the idea of qualia is mistaken, and that we can fully understand consciousness without it. Dennett goes to some length to disabuse us of the belief that our experiences have such properties (1986). But, as I have argued, the idea of qualia is essential to the idea of conscious experience – we cannot think of any experience without thinking of its having an irreducible *quality*. It strikes us that what it is like to experience things is fundamental, i.e., it is something that is constitutive of us as phenomenal subjects, and not something that can be explained away. What distinguishes experiences from other phenomena we study is that they have these peculiar properties or qualities.

As already noted (section 4.3.1, n. 31), in one of his arguments for the viability of heterophenomenology Dennett imagines the Earth being visited by an advanced civilisation of Martians whose scientists endeavour to study us, including our consciousness. These Martians are simply thought of as phenomenal zombies, and not as physically and behaviourally indistinguishable from us of course. Thus, it is only through the third-person viewpoint that we and the Martians can relate to one another. Open to the Martians, then, is the study of our physiology and behaviour broadly construed to include first-person reports of our phenomenal states. Therefore, the Martians are free to employ a heterophenomenological approach to the study of human consciousness. Thus, according to Dennett, they are able to develop a complete, or completable, theory of human consciousness in neuroscientific terms.

But, as zombies they would not have arrived on Earth equipped with the concept of consciousness. That is because there would have been no use for such a concept in their world. Were one of the Martians to exhibit pain behaviour, for example, they would have no need to state how this 'experience' felt, since there would have been nothing that it is like to be in pain for them. They would have no terms equivalent to our 'feel' or 'seem'. Consequently, *ex hypothesi* on encountering these terms on Earth, they could comfortably define them in terms of the physiological and behavioural properties posited by neuroscience – this would be the only way, after all, that they could make sense of such terms and thereby be able to *use* them. The idea that the report 'I feel a sharp pain', for example, alludes to an experience also understood in terms of its phenomenological quality would be completely passed over by them. Consequently, they could not grasp the concept of being phenomenally conscious as such. So, for example, these Martians would

be unable to distinguish between the meanings of the statements 'the tomato seems red to him' and 'he sees (senses) the tomato as red'.

A similar point is made by Todd Moody, who argues that within a community of zombies, who are otherwise like us, certain psychological concepts would be absent from their language.[61] Zombies could not originate such psychological terms as 'consciousness', 'seeing' and 'dream' since such mental states do not exist for them. In reply to Moody Owen Flanagan and Thomas Polger argue that while it is highly unlikely that zombies would develop these terms it is not logically impossible for them to do so. For example, they imagine how occasionally zombies walk into trees and these and similar events lead their compatriot zombies to shout the warning "Watch out!" This kind of talk would allow the zombies to develop the concept of seeing, roughly 'understood' by them to refer to one's photoreceptors being oriented in the right direction (Flanagan and Polger, 316). Flanagan and Polger argue that similar stories could be given for the origination of other psychological terms in the zombie speech community. But all these zombie terms ultimately allude to aspects of behaviour. The zombies would have no inclination to think that there is something that it is like to see, dream, or bump one's head against a tree. It would still be impossible for them to wonder what is it like to see a ripe tomato as red. The idea of experiences having a qualitative character could never occur to them. So, the term 'pain', for example, could only be used by them to allude to certain behavioural dispositions; there would be no notion of the hurtfulness of pain, no question of this *feeling* different from being tickled by a feather, say.

So, Dennett's Martians would end up 'grasping' *our* concept of consciousness in naturalistic terms. On the other hand, we have a complete grasp of the concept, i.e., both

in naturalistic and phenomenological terms – this latter understanding would be entirely missed by the Martians. From their so-called perspective a heterophenomenological approach to the study of consciousness, therefore, would offer the promise of a full and satisfying theory of consciousness. But from our perspective this approach does not accommodate our phenomenological understanding of consciousness, given that this understanding plays no role in the resultant neuroscientific theory. Dennett does not see such a lack of accommodation because he denies the existence of qualia.

While heterophenomenology cannot accommodate our phenomenological understanding of consciousness I do not count this as a failure. A heterophenomenological approach to the study of consciousness allows us to relate *talk* of how things seem to us to neuroscience. But because qualia are the properties in virtue of which we construct our theories of the world, including consciousness, they cannot themselves be posits of our theories. Thus, by thinking of qualia as epistemically originating properties in this way we can overcome the worry that neuroscience says nothing about qualia. There is no sense in which it could do so. This does not point to some limitation of neuroscience. The neuroscientist can explain consciousness as we understand it in terms of physiology and behaviour.

5.4 Qualia as Epistemically Originating Properties

I have argued that what is peculiar about qualia is that they are epistemically originating properties. Below, I want to explore in a little more detail what this construal of qualia amounts to. An epistemically originating property is a property realised by conscious creatures in virtue of which they apprehend some aspect of the world. If a person, S,

---

[61] See Moody 1994, 196-200.

apprehends something as yellow, i.e., something seems yellow to S, then S realises a yellow quale. This yellow quale is epistemically originating because only in virtue of realising it does S apprehend something in the world as yellow. However, this explanation appears to be circular – S apprehends some aspect of the world, e.g., yellow, because S realises an epistemically originating property, and S realises an epistemically originating property because it apprehends some aspect of the world. What work does the notion of an epistemically originating property do? Such properties have the look of occult properties.

But we can avoid this empty way of understanding epistemically originating properties if we think of them as the properties by which the subject is realised. These properties seem occult so long as we think of the subject or self as existing independently of them, so that if the subject does not realise any qualia it would nonetheless still exist; in other words, they seem occult so long as we think of them as contingent properties of the subject. However, if we think of qualia as properties that constitute the subject or self, i.e., as properties necessary for there to be a self, then we can offer a non-circular explanation of them. In explaining that S apprehends something as yellow because S realises a yellow quale, we are led to ask: why does S realise a yellow quale? Instead of replying because S apprehends something as yellow, we can say that S realises a yellow quale in the sense that S is partly constituted by it. Being S is in part to realise qualia generally. Being S, i.e., being a subject or self, is to apprehend aspects of the world. It is important to note here that the claim is not that realising qualia is a *sufficient* condition for being a self, only that it is a necessary condition.

Crucially, as noted earlier, there is no distinction to be made between an experience having a particular quale and an experience *seeming* to have this quale (see section 3.2). Qualia as ways things seem to us cannot themselves seem some way to us; they are the appearances themselves.[62] In this respect qualia are transparent to us. When we try to apprehend them in this way we *see right through* them so to speak to the external properties of the things the experience is of. Understanding qualia as constitutive of us as phenomenal subjects in this way explains their perspectival nature, pointed out so perspicuously by Nagel. Thinking of qualia as properties by which we apprehend things in the world explains their essential connection to a single point of view.[63] Qualia, one might say, are epistemological properties through and through – they are what characterise the subject in contradistinction to the world the subject apprehends and thereby comes to know about.

We can think of realising qualia as being phenomenally conscious, so that some being is, or realises, a self only if it is phenomenally conscious. By this measure, a zombie cannot realise or be a self. And indeed a zombie is like a chair or some other inanimate object, that is, there would be nothing that it is like to be such a thing. And in general a being is phenomenally conscious if and only if it has a point of view. Without a point of view there is no world for a being – it does not stand in contrast to the rest of the world,

---

[62] Again, this basic point is made by Saul Kripke in his *Naming and Neccesity*, where he states that "in the case of mental phenomena there is no 'appearance' beyond the mental phenomenon itself" (1980, 154).

[63] This general way of thinking of qualia is urged by Fred Dretske and Mark Rowlands for example. Rowlands writes: "What it is like is an aspect of conscious experience that exists only in the directing of such experience towards a non-phenomenal object. It is not itself an object of such experience" (2001, p. 149). And Dretske writes: "Conscious mental states – experiences, in particular – are states that we are conscious *with*, not states we are conscious *of*" (1995, pp. 100-101).

so to speak. Nothing exists *for* a zombie. So, a zombie's walking into a tree, to borrow Flanagan and Polger's example, is neither a good nor a bad thing since it feels no pain.

Now, one might argue that the zombie could evolve unthinking or moronic behaviour aimed at 'avoiding' such events by the process of natural selection. Walking into trees or any other solid object, after all, is bad at least in the sense that it is not conducive to survival. Hence, zombies would have an interest in this minimal sense. But *for* this zombie nothing happens when it walks into a tree. Its walking into a tree does not constitute an event for it. This minimal interest in survival is external to it, that is, it would not take it on as *its* interest – to do so requires it to experience pain, and *a fortiori* pleasure. A zombie would be like a plant in this respect – an equally insentient being – which likewise has an 'interest' in survival but no interests of its own.

However, one could perhaps argue that being a self is not essentially being phenomenally conscious. Because a zombie has a body it is a self in the sense of being an actor in the world. A zombie's world exists as an arena for its actions. Consequently, zombies can at least realise intentional states, i.e., have propositional attitudes, such as beliefs about the world. These beliefs would have content insofar as they guide the zombie's actions. It is this assumption that allows Dennett, for example, to assume that his zombie Martians can adopt an intentional stance. But, the difficulty is that in the case of a supposed zombie no *one* is there. As a being without a *particular* point of view there would be no contrast beween the zombie and the world. On the other hand, I exist *in* the world as a self because I distinguish myself from the world – that is what it is to be a self. For a zombie no such distinction exists – there is no difference between the world and the zombie. The concepts of the self and of the world are intimately connected in this way, as

implied by their logical relation, i.e., the self versus the not-self. It is this connection between self and world that is key. In this sense, we could attribute intentional states to a zombie, e.g., desires and interests, in essentially the same way that we can attribute such states to a company, say, or to a plant under certain circumstances perhaps. However, being able to *attribute* intentional states to some thing does not by itself entail this thing is or realises a self.

5.5 Dissolving the Problem of Consciousness

We started with the problem of consciousness. How do we accommodate subjectivity in the natural world governed by the laws of physics? The phenomenological qualities of our experiences, or qualia, that define this subjectivity seem to bear no relations to the world as we understand it in terms of our sciences. Consciousness is essentially characterised in terms of qualia. What we mean when we say someone is phenomenally conscious is that they realise these qualia, that is, they have experiences we define in terms of these properties. The phenomenon of consciousness, then, seems to stand outside of our scientific investigations – none of our scientific theories can help us understand what consciousness is. Consciousness seems to be unnaturalisable.

In the introduction I noted that Valerie Hardcastle identifies two opposing camps concerning the problem of consciousness. One camp, the 'naturalists' as she calls them, which includes herself, comprises those who think that consciousness is in principle naturalisable. They claim that there is no reason to suppose that we cannot eventually arrive at a satisfactory theory of consciousness. The other camp, the 'sceptics', comprises those who are sceptical of this possibility. According to Hardcastle there is an

unbridgeable gap in attitude between these camps. I said that one of my overall aims is to show how *pace* Hardcastle this gap is bridgeable and to encourage a way of thinking about the concept of consciousness that dissolves the problem of consciousness. We are now in a position to summarise how this can be done.

We have seen that there are four ways of understanding and dealing with the problem that relate directly to Hardcastle's analysis. For those who assume consciousness can be explained by our sciences, like Hardcastle, the problem then comes down to either (1) explaining how each quale is identical with some physiological state or (2) explaining how there is nothing more to understanding consciousness than what our sciences tell us about it. On the other side, the sceptics suppose that the nature of qualia is such that there is no possibility of naturalising consciousness. This sceptical approach divides into two, namely, arguing either that (3) our sciences can never provide us with a theory of consciousness, or (4) our sciences need to be reformed in order to make such a theory possible. (1) essentially describes the position of Flanagan and Papineau, for example. (2) is the position adopted by Dennett, who dismisses the concept of qualia as irredeemably confused. (3) describes McGinn's general view, where he argues that the non-spatiality of consciousness precludes us from ever constructing a theory of consciousness. And (4) fits with the view of Chalmers, whose recommended reform concerns thinking of conscious properties as basic, i.e., on a par with such basic physical properties as length and having a mass. Importantly, I have argued that ultimately none of these approaches are by themselves satisfactory.

I began the discussion proper, in chapter 1, by distinguishing between two opposing philosophical attitudes, which I called 'naturalistic' and 'non-naturalistic'. These attitudes

correspond to what Hardcastle calls the 'naturalists' and 'sceptics' respectively. The naturalistic attitude I defined as one that takes philosophy to be continuous with science. The ultimate mark of this attitude is the belief that there is no higher tribunal regarding our judgments about reality than our senses. The non-naturalistic attitude, on the other hand, is defined by the belief that our intuitions are equally authoritative vis-a-vis our judgments about reality. I argued that adopting a naturalistic attitude towards philosophical problems quite generally is better – it avoids the extravagant metaphysical claims that a non-naturalistic attitude often entails, which result from their not being falsifiable through the senses, i.e., of being constrained in any real sense. Moreover, adopting this attitude is helpful in overcoming the problem of consciousness specifically. How it helps has been a central theme of the dissertation. And in this chapter I argued that to the extent that we endeavour to provide a scientific explanation of consciousness this cannot be directly informed by how we intuitively understand it, i.e., in terms of qualia.

In chapter 2 I considered arguments presented by Nagel, the early Jackson, and Chalmers that aim to show that consciousness is non-physical so that science as it is presently practised or understood either cannot solve the problem of consciousness or must be reformed radically to do so. These positions generally concur either with (3) or (4). I showed, by means of well-known objections to them, that none of these arguments are successful. The arguments, I pointed out, betray a non-naturalistic approach to the problem. In this respect these views are characterised by the assumption that our beliefs concerning the phenomenological qualities of experience, i.e., qualia, are in principle as justifiably true as those concerning ordinary physical properties. I questioned this

assumption, and later challenged it outright in terms of claiming that qualia are unindividuable.

It is in chapter 3 that I urged that we think of qualia as unindividuable. The nature of qualia is such that there can be no criterion by which we can objectively determine if two qualia are distinct from or identical with each other. To deny their *un*individuability is to abandon thinking of qualia as such. The unindividuability of qualia, I argued, follows from an understanding of them as properties *by which* we apprehend things in the world rather than as properties we apprehend in the world. Ultimately I came to refer to qualia in this sense as epistemically originating properties. This way of understanding qualia enables us to make sense of their perpectival nature, i.e., their essential connection to a particular point of view. In this chapter I observed that Dennett likewise thinks of qualia as unindividuable, although he concludes from this that the notion of qualia is confused and should be abandoned entirely. Here I marked a disagreement with his conclusion.

In chapter 4 I looked at the worry that because consciousness is understood in terms of qualia any scientific theory of consciousness misses the target, that is, it cannot explain what is essential to consciousness and so it is useless. In reply, I argued that while how we think of consciousness must include this phenomenological aspect of it, this way of thinking of consciousness only exists because we also think of it in terms of physiology and behaviour. These two ways of thinking of consciousness are inseparable. Accordingly, though we cannot identify qualia with certain physiological states given that qualia are unindividuable, we cannot think of consciousness solely in phenomenological terms so that we might suppose that science has nothing to say about it.

So in reply to Hardcastle's claim that the gap in attitude she describes is unbridgeable, that is, there seems no way to persuade members from one camp to adopt the attitude of the other, we now see how this is not the case. The non-naturalistic attitude of the sceptics is seriously disadvantaged by the fact that their appeal to intuition means their metaphysical claims are ultimately unfalsifiable. Consequently, there is no constraint on the claims they make. Chalmers speculates that conscious properties are ubiquitous so that everything realises them, but unfortunately the protophenomenal properties he postulates are in principle unobservable. McGinn thinks that non-spatial properties that constitute consciousness, which again are in principle unobservable, can be incorporated into science by some sort of *a priori* method. How this is at all possible is unclear. All he tells us is that it requires a degree of intelligence beyond our reach. The naturalist, on the other hand, has at his disposal a tangible means of grounding our theories about consciousness. This concerns supposing that there is no higher authority regarding our judgments about the world than the tribunal of our senses. Having this anchor so to speak allows us to measure the ultimate plausibility of any theories we might devise. This is an overwhelming advantage for the naturalist.

So, how exactly does the naturalistic attitude help us dissolve the problem of consciousness? The problem is motivated by the assumption that qualia are individuable. By thinking of qualia as individuable one is led to ask why we cannot understand *them* in terms of other natural phenomena, that is, why we cannot naturalise them. But qualia, I have argued, are unindividuable and understanding them to be so dispels the problem – it makes the problem nonsensical. That said, it is difficult to think of qualia as being both unindividuable and real. If we cannot ultimately determine whether one quale is distinct

from or identical with another quale, then it is hard to imagine how qualia can be thought of as properties at all. A major part of my project has been to reconcile these claims, namely, that qualia are unindividuable and that they are real, i.e., they exist.

In reply, I have argued that qualia are properties that determine our particular point of view, that is, as phenomenal subjects we are constituted by qualia. This is to think of qualia as properties by which we apprehend things in the world as epistemically originating properties. Without them we can apprehend nothing in the world, i.e., no world would exist for us. Further, they cannot be distinguished from one another because they are the very properties in virtue of which we are able to make such distinctions vis-a-vis every property that we posit in our theories about the world. Likely this construal of qualia as epistemically originating properties will strike some as dubious, as a sleight of hand. One might suspect that it is an attempt to wallpaper over the problem of consciousness rather than facing up to it: it is to treat qualia as occult properties – we do not have to explain them because they are the properties by which we are able to explain everything else. Thus, qualia are *intrinsically mysterious* properties.

This worry, I argued, again results from presupposing that there is a viewpoint from which everything is graspable, that is, a perspective from which we can make judgments about anything at all including those properties constitutive of our powers of judgment, as if we have limitless powers in this regard. In other words, it results from adopting a non-naturalistic attitude. But as finite creatures our theories about the world are constructed through our senses. The world understood in terms of our theories is always apprehended from some particular point of view. As Quine points out: "Our talk of external things, our very notion of things, is just a conceptual apparatus that helps us to foresee and control

the triggerings of our sensory receptors in light of previous triggerings of our receptors" (1981, 1). We have no other means of judging ultimately what is and what is not the case than our senses, no extrasensory perception. The triggerings of our sensory receptors are all we have to go by. In virtue of such triggerings you and I exist, that is, somehow these triggerings lead to a contrast between the world and its apprehension, manifested as the self. How does this contrast arise? We can tell various stories about how the brain represents these triggerings as something in the world, or about how these triggerings as input are related to other inputs as well as to various behavioural and physiological outputs, and so forth. But if one were to go on to ask why these representations, functional roles, etc. *feel* some way rather than like nothing at all, our only reply can be that that is how it is. We reach the limit of our explanations. Does this amount to a mystery? Yes, it amounts to the same kind of mystery expressed by the question 'why is there something rather than nothing at all?' And such questions have no answers.

How do we understand consciousness as a phenomenon in the world governed by the laws of physics? Insofar as consciousness is thought of in terms of physiology and behaviour there is no difficulty in explaining it scientifically. As far as consciousness understood in phenomenological terms is concerned, there is no 'hard' problem of explaining it scientifically because the properties that define this understanding, i.e., qualia, are epistemically originating and hence unindividuable. They are not entities which we can quantify over and therefore subsume into our theories. When, for example, Chalmers asks "[w]hy should physical processing give rise to a rich inner life at all?" (1995, 201), we answer that the 'life' he alludes to is not inner, it is not something we each grasp in some privileged manner by glancing inwards. The subject does not

apprehend the phenomenological qualities of experience, rather they are constitutive of the subject – qualia are the manifestation of a subject. In this sense each experience is an episode *of* the subject. It is not *hard* to incorporate consciousness understood in terms of qualia, i.e., phenomenologically, into our scientific theories, it is impossible. But understanding why it is impossible, namely, for the reasons I have given, we can see that there is nothing science fails to explain.

In conclusion, the claims I have made are obviously not without difficulties. For example, more needs to be said about the idea of an epistemically originating property. The phrase is slightly misleading in the sense that it might suggest to some that such properties are the grounds for our knowledge, that is, it alludes to the sort of foundationalism Wilfrid Sellars famously railed against in terms of the Myth of the Given.[64] The idea which one might be misled to think of is roughly that our beliefs about natural phenomena are ultimately founded on a set of basic beliefs about qualia that are non-inferentially acquired; moreover, having these basic beliefs about qualia is not dependent on having any other beliefs. This is not what I mean by epistemically originating properties. They are not properties which in virtue of realising we can acquire some set of foundational beliefs. Rather, the idea is that they are properties in virtue of which we are able to apprehend the world, that is, they are the grounds for our acquiring beliefs about the world, i.e., natural phenomena, very generally. This difference is important and may be worth exploring in its own right.

Lastly, throughout I have remained largely unconcerned with the truth of physicalism. That is, from the beginning I have stated that my principal concern is to understand how consciousness is naturalisable, and not to defend physicalism per se. But this lack of

concern with the truth of physicalism might seem to be intolerable with respect to my construal of qualia in terms of causation. Consider specifically Joseph Levine's observation that while we cannot understand how experiences picked out by their qualia can be identical with some physical states, i.e., the problem of the explanatory gap, this fact does not entail such identities are false. But more importantly, the truth of physicalism seems the only way to allow us to understand how mental phenomena in general are causally efficacious. This is to assume epiphenomenalism is incompatible with physicalism. As Levine puts it: "It seems overwhelmingly obvious that mental phenomena are both causes and effects of non-mental, physical phenomena."[65] We want to say that physicalism just has to be true but damned if we know how it is. Given the unindividuability of qualia I concluded that psychophysical identities concerning them cannot be asserted, and so I deny that we can hold such identities as true.[66] But, one might worry that so long as qualia are taken to be causally efficacious they must be thought of as physical. Yet I have effectively denied that we can *understand* qualia as physical properties. This would make physicalism an incoherent doctrine. If physicalism is true then qualia, as real properties, must be causally efficacious. But if qualia cannot be identified with any physical properties it seems we cannot know whether they are causally efficacious or not. Therefore, we *cannot* determine whether physicalism is true or false.

In reply, I would say that the unindividuability of qualia and the consequent impossibility of identifying them with anything does not indicate the incoherence of

---

[64] See Sellars 1963, 164-167.
[65] Levine 2001, 5.

physicalism *simpliciter*. Rather, I think that this difficulty points to physicalism not being entirely coherent. Indeed, here I would point to our inability to define the term 'physical'. This difficulty is analogous to that of accommodating irrational numbers on the continuum. If the continuum is thought of as a sequence of contiguous points each expressible as a fraction, i.e., as a rational number, there seems to be no place for irrational numbers like π and $\sqrt{2}$, which cannot be reduced to fractions, hence their irrationality. No matter how many decimal places one calculates the value of π, for example, one only arrives at an approximation to it. There is no point on the continuum that can correspond with π's value. This suggests that not every number is on the continuum, which is a paradoxical conclusion. Some numbers are larger than π and some are smaller and as such it has a place, so to speak, among numbers quite generally. In other words, the idea of the continuum seems *incoherent*. The generally agreed on way of making the idea of the continuum coherent is to accept Richard Dedekind's suggestion. Dedekind urges that irrational numbers be understood as cuts in the continuum. A cut is thought of as existing between contiguous points on the continuum, i.e., as a partition betwen rational number sequences, rather than as a gap in it, which contradicts the very idea of continuity of course.[67]

The approach that this Dedekindian solution to the problem of accommodating irrational numbers on the continuum adopts is entirely naturalistic in tenor. Mathematicians did not reject the idea of the continuum as incoherent, rather they

---

[66] And the unindividuability of qualia rules out our being able to assert their identity with anything at all, including functional roles and behavioural dispositions as well as physical states.

[67] More precisely, an irrational number is thought of as the partition of two disjoint subsequences such that all the members of one set of a rational number sequence are

adjusted their conceptions of numbers to make the idea as coherent as possible. I would urge the same approach with respect to physicalism. As a scientific hypothesis I take physicalism to be true. And insofar as understanding qualia as unindividuable makes it impossible to see how it is true, one should try to conceive of qualia in a way that makes them compatible with the truth of physicalism. Therefore, to paraphrase Jaegwon Kim, we should accept physicalism as near enough true in this sense.[68] That is what I have endeavoured to do. This naturalistic approach to understanding consciousness, I hope to have shown, is the best one to take.

---

larger than all the members of the other set, e.g., $\sqrt{2}$, is defined as the following ordered pair of sets $<\{x: x^2 > 2\}, \{x: x^2 < 2\}>$ (see Borowski and Borwein, 145).

[68] See Kim 2005, 174.

Bibliography

Ambrose, Alice. (1979). *Wittgenstein's Lectures, Cambridge, 1932-1935: From the Notes of Alice Ambrose and Margaret Macdonald.* NY: Prometheus Books.

Armstrong, David H. (1981). "What is consciousness?" *Philosophy of Mind: A Guide and Anthology*, (2004), Heil, John, ed., Oxford: Oxford Univ. Press., pp. 607-616.

Austin, J.L. (1962). *Sense and Sensibilia.* Oxford: Clarendon Press.

Block, Ned. (1994). "Consciousness." *A Companion to the Philosophy of Mind.* Guttenplan, S., ed., Oxford: Blackwell Pub., pp. 575-583.

Borowski, E.J., and Borwein, J.M., eds. (1989). *Collins Dictionary of Mathematics.* NY: Harper Collins.

Boyer, Carl B. (1949). *The History of Calculus and Its Conceptual Development.* NY: Dover Pub. Inc.

Brueckner, Anthony, and Beroukhim, E. Alex. (2003). "McGinn on Cosciousness and the Mind-Body Problem." *Consciousness: The New Philosophical Perspective.* Smith, Quentin and Jokic, Aleksandar, eds. Oxford: Clarendon Press, pp. 396-406.

Carnap, Rudolph. (1934/1995) *The Unity of Science*. Bristol: Thoemmes Press.

Carpintero, Manuel García- (2003). "Qualia that It Is Right to Quine." *Philosophy and Phenomenological Research*, v. LXII, No. 2: 357-377.

Carruthers, Peter.

(2000). *Phenomenal Consciousness*. Cambridge: Cambridge Univ. Press

(2001). <www.swif.uniba.it/lei/mind/forums/carruthers4.htm>.

(2004). "Reductive Explanation and the 'Explanatory Gap'." *Canadian Journal of Philosophy*, 34: 153-174.

Chalmers, David.

(1995). "Facing up to the Problem of Consciousness." *Journal of Consciousness Studies* 2, No. 3: 200-219.

(1996). *The Conscious Mind*. Oxford: Oxford Univ. Press.

Churchland, Paul.
(1985). "Reduction, Qualia, and the Direct Introspection of Brain States." *Journal of Philosophy* 82, No. 1: 8-28.
(1989). *A Neurocomputational Perspective*. Cambridge, MA: MIT Press.
(1998). "The Rediscovery of Light." *On The Contrary*. Cambridge, MA: MIT Press, 123-141.
(1998). "Knowing Qualia: A Reply to Jackson." *On the Contrary*. Cambridge, MA: MIT Press, pp. 143-157.

Coulter, Jeff, and Sharrock, Wes. (2007). "Chapter Four: Consciousness: the Last Mystery?" *Brain, Mind, and Human Behaviour in Contemporary Cognitive Science: Critical Assessments of the Philosophy of Psychology*. Lewiston, NY: Edwin Mellen Press, pp. 83-136.

Damasio, Antonio. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*, London: Penguin Books.

Davidson, Donald. (1970). "Mental Events." *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, pp. 207-227.

Dennett, Daniel.,
(1988). "Quining Qualia." *Philosophy of Mind: Classical and Contemporary Readings*. D. Chalmers, ed., Oxford: Oxford Univ. Press, 2002, pp. 226-246.
(1987). *The Intentional Stance*. Cambridge, MA: MIT Press.
(1991). *Consciousness Explained*. Boston, MA: Little, Brown Co.
(2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press.

Descartes, René. (1985). *The Philosophical Writings of Descartes,* vol. 1. Cottingham, J., Stoothoff, R. and Murdoch, D., trans. Cambridge: Cambridge Univ. Press.

Dretske, Fred. (1997). *Naturalizing The Mind*. Cambridge, MA: MIT Press.

Flanagan, Owen. (1992). *Consciousness Reconsidered*. Cambridge, MA: MIT Press.

Flanagan, Owen and Polger, Thomas. (1995). "Zombies and the Function of Consciousness." *Journal of Consciousness Studies* 2, No. 4: 313-321.

Ford, Kenneth W. (2004). *The Quantum World: Quantum Physics for Everyone*. Cambridge, MA: Harvard Univ. Press.

Frege, Gottlob.
(1980). *The Foundations of Arithmetic*. 2nd edition, Austin, J.L., trans., Evanston, IL: Northwestern Univ. Press.
(1997). "*On* Sinn *and* Bedeutung." *The Frege Reader*. Beaney, Michael, ed., Black, Max, trans., Oxford: Blackwell Pub.

Garber, Daniel. (1992). *Descartes' Metaphysical Physics*. Chicago: Univ. of Chicago Press.

Goodman, Nelson. (1983). *Fact, Fiction and Forecast*, 4th edition. Cambridge, MA: Harvard Univ. Press.

Hankinson-Nelson, Lyn, and Nelson, Jack. (2000). *On Quine*, Belmont, CA: Wadsworth Inc.

Hardcastle, Valerie.
(1996a). *How to Build a Theory in Cognitive Science*. Albany, NY: SUNY Press
(1996b). "The Why of Consciousness: A Non-Issue for Materialists." *Journal of Consciousness Studies* 3, No. 1: 7-13.

Hardin, Clyde.
(1987). "Qualia and Materialism: Closing the Explanatory Gap." *Philosophy and Phenomenological Research* Vol. XLVIII, No. 2.
(1988). *Color for Philosophers*. Indianapolis, IN: Hackett Pub. Co.

Hill, Christopher S. (1991). *Sensations: A Defense of Type Materialism*. Cambridge: Cambridge Univ. Press.

Honderich, Ted. (2004). *On Consciousness*. Pittsburgh, PA: University of Pittsburgh Press.

Humphrey, Nicholas. (1992). *A History of the Mind: Evolution and the Birth of Consciousness*. NY: Springer Verlag.

James, William.
(1890/1950). *The Principles of Psychology* v.1, New York: Henry Holt & Co. Reprinted in 1950, NY: Dover.
(1907/1995). *Pragmatism*. NY: Dover

Jackson, Frank.,
(1982). "Epiphenomenal Qualia." *Philosophical Quarterly* 32: 127-136.
(1986). "What Mary Didn't Know." *Journal of Philosophy* 83: 291-295.

Jesseph, Douglas M. (1992). *De Motu and The Analyst*. Dordrecht, Netherlands: Kluwer Academic Pub.

Kim, Jaegwon.
(1993). *Supervenience and Mind*. Cambridge: Cambridge Univ. Press.
(1994). "Supervenience." *A Companion to the Philosophy of Mind*. S. Guttenplan, ed., Oxford: Blackwell Pub., pp. 575-583.
(2005). *Physicalism or Something Near Enough*. Princeton, NJ: Princeton Univ. Press.

Kriegel, Uriah. (2005). "Naturalizing Subjective Character." *Philosophy and Phenomenological Research*, vol. 71, 1: 23-57.

Kripke, Saul. (1980). *Naming and Necessity*. Oxford: Blackwell Pub.

Levine, Joseph.
(1983). "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64:354-61.
(2001a). *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford Univ. Press.
(2001b). < www.swif.uniba.it/lei/mind/forums/levine.htm>.

Lewis, David. (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge Univ. Press.

Loar, Brian. (1997). "Phenomenal States." *Philosophy of Mind: Classic and Contemporary Readings*. Chalmers, D.J., ed., NY: Oxford Univ. Press (2002).

Lycan, William. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.

McGinn, Colin.
(1989). "Can We Solve the Mind-Body Problem?" *Mind* 98: 349-66.
(1991). *The Problem Of Consciousness: Essays Towards a Resolution*. Oxford: Blackwell Pub.
(1999). *The Mysterious Flame*. NY: Basic Books.

Malcolm, Norman. (1958) "Knowledge of Other Minds." *The Journal of Philosophy LV*, 23: 969-78.

Melnyk, Andrew.
(2001). "Physicalism Unfalsified." *Physicalism and Its Discontents*. Gillett, Carl, and Loewer, Barry, eds. Cambridge: Cambridge Univ. Press.
(2003). *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge: Cambridge Univ. Press.

Misak, C.J. (1995). *Verificationism: Its History and Prospects*. London: Routledge.

Moody, Todd. (1994). "Conversations with Zombies." *Journal of Consciousness Studies* 1: 196-200.

Nagasawa, Yujin. (2003). "Thomas vs. Thomas: A New Approach to Nagel's Bat Argument." *Inquiry* 46: 377-394.

Nagel, Ernest. (1961). *The Structure of Science*, NY: Harcourt, Brace & World Inc.

Nagel, Thomas.,
(1974). "What Is It Like to Be a Bat?" *The Philosophical Review* LXXXIII, 4: 435-50.
(1986). *The View From Nowhere*. Oxford: Oxford Univ. Press.

Papineau, David. (1993). *Philosophical Naturalism*. Oxford: Blackwell.

Place, U.T. (1956). "Is Consciousness a Brain Process?" *British Journal of Psychology* 47:44-50.

Putnam, Hilary. (1981). *Reason, Truth and History*. Cambridge: Cambridge Univ. Press.

Quine, W.V.O.,
(1960). *Word and Object*. Cambridge, MA: Harvard Univ. Press.
(1979). "Facts of the Matter." *Essays on the Philosophy of Quine*. Shahan, Robert, and Swoyer, Chris, eds., Oklahoma: Harvester Press.
(1981). *Theories and Things*. Cambridge, MA: Harvard Univ. Press.
(1992). *Pursuit of Truth*. revised edition, Cambridge, MA: Harvard Univ. Press.
(2004). *Quintessence: Basic Readings from the Philosophy of W.V.Quine*, Gibson, Roger F., ed., Cambridge, MA: Harvard Univ. Press.

Rosenthal, David. (1986). "Two Conceptions of Consciousness." *Philosophical Studies*, 49: 329-59.

Ross, Don. (1993). "Quining Qualia Quine's Way." *Dialogue* **32**: 439-59

Rowlands, Mark. (2001). *The Nature of Consciousness*. Cambridge: Cambridge Univ. Press

Ryle, Gilbert. (1949). *The Concept of Mind*. London: Penguin.

Seager, William.
(1999). *Theories of Consciousness*. London: Routledge.
(2000). "Real Patterns and Surface Metaphysics." *Dennett's Philosophy: A Comprehensive Assessment*. Ross, D., Brook, A., and Thompson, D., eds., Cambridge, MA: MIT Press, 95-129.

Searle, John. (1994). *The Rediscovery of Mind*. Cambridge. MA: MIT Press.

Sellars, Wilfrid. (1963). *Science, Perception and Reality*. London: Routledge & Kegan Paul.

Shoemaker, Sydney. (1996). "Intrasubjective/Intersubjective." *The First-Person Perspective and Other Essays*. Cambridge: Cambridge Univ. Press, 141-154.

Smith, Barry, and Woodruff Smith, David. (1995). "Introduction." *The Cambridge Companion to Husserl*. Cambridge: Cambridge Univ. Press.

Smolin, Lee. (2006). *The Trouble with Physics*. Boston, MA: Mariner Books.

Strawson, Galen. (1994). *Mental Reality*. Cambridge, MA: MIT Press.

Strawson, P.F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. Garden City, NY: Anchor Books.

Tye, Michael. (1995). *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

Wider, Kathleen. (1990). "'Overtones of Solipsism in Thomas Nagel's 'What is it like to be a bat?' and the *View from Nowhere*." *Philosophy and Phenomenological Research*, v. 50, no. 3: 481-499.

Williams, Bernard,

(1973). *Problems of the Self: Philosophical Papers 1956-1972*. Cambridge: Cambridge Univ. Press.

(2006). *Philosophy as a Humanistic Discipline*. Moore, A.W., ed., Princeton, NJ: Princeton Univ. Press.

Wittgenstein, Ludwig.

(1922). *Tractatus Logico-Philosophicus*. London: Routledge.

(1958). *Philosophical Investigations*. Anscombe, G.E.M., trans., Eaglewood Cliffs, NJ: Prentice-Hall.