Final version: 3 June 2016

The Possibility of Materialism

Mike Holliday

1: Introduction

Is a materialist account of conscious experience even *possible*? David Chalmers famously answered 'No', setting out an argument based upon a distinction between the easy problems of explaining our mental processes and the hard problem of accounting for experience.

Of course it is not unusual to hear the opinion put forward that there cannot be an explanation of consciousness in terms of physics, or the brain, or information processing, because experiences such as smelling a flower or suffering an intense pain are completely different in nature from electrons, or neurons, or bits and bytes. That's the sort of thing which people were saying well before David Chalmers stood up at the Tucson conference in 1994 and started talking about the 'hard problem' – there were *always* those who were dubious about materialist explanations of consciousness. Others, however, believed that the scientific evidence increasingly pointed to experience being something going on in the brain – that experience was something the brain just *did* – and that the qualitative aspects of experience were therefore most likely to be some form of higher-level property of the brain, rather than *intrinsic* properties that were left undetermined by the physical goings-on inside the human skull.

The significance of Chalmers' intervention in this ongoing debate, as set out in his talk at Tucson and in his subsequent paper 'Facing Up to the Problem of Consciousness' (1995, hereafter 'Facing Up'),[1] was that he claimed to be demonstrating that the materialists had to be wrong: that it was simply impossible – not implausible, or difficult to demonstrate, but *impossible* – for there to be an explanation of conscious experience in wholly physical or functional terms.[2] We were, Chalmers insisted, looking in completely the wrong place for an understanding of experience and of how we might scientifically theorize it. But even after 20 years, this answer still has something strange about it: notwithstanding all the scientific evidence that relates our mental states to neurophysiological

---

[1] 'Facing Up' is now included in Chalmers' collection of articles *The Character of Consciousness* (2010). For the sake of convenience, I reference wherever possible to that volume rather than to the original articles. All references herein are to works by Chalmers, unless otherwise indicated.

[2] Chalmers takes a functional analysis to be the only type of physical analysis of a phenomenal property that is 'even remotely tenable' (1996, p. 104). Others, such as Ned Block (2015), argue that functional and physical explanations are orthogonal to each other – and that functionalism and (at least) some types of physicalism are mutually exclusive. I take no position here on this issue, and will normally use the phrase 'physical *or* functional'.

activity, the one account of conscious experience which is firmly ruled out of bounds as even a *possibility* is a materialist account.

Now if someone puts forward an argument with the aim of proving that X is not possible, there are two different ways of contesting that argument. One is to produce an account which demonstrates that X actually *is* the case, and the second is to show that the argument put forward against X's possibility is flawed. In what follows I will be solely interested in taking the second approach, by means of a critical examination of the argument against materialism as Chalmers sets it out in his articles and books.

However, anyone hoping to show that Chalmers' argument is mistaken faces a number of obstacles. The first difficulty is obtaining a clear understanding of what the argument actually *is*. That this is no easy matter can be seen from a perusal of the invited responses to 'Facing Up', and Chalmers' reply (1997) to those responses. Most of the commentators felt that they had met Chalmers' objection to materialist explanations of experience, but he was adamant that they had not. Time does not seem to have made matters any easier. Some 15 years later, in responding to a paper by Glenn Carruthers and Elizabeth Schier which had criticised his formulation of the 'hard problem', Chalmers commented that: 'I'd thought my [earlier] articles were clear about these things, but obviously they aren't as clear as they could have been, so I'll try to be really explicit' (2012a) – and yet his reply still left Carruthers and Schier puzzled.

A second difficulty arises because in *The Conscious Mind* (1996) Chalmers makes the case against materialism in a rather different way from that in 'Facing Up'. The book-length treatment proceeds by way of a lengthy, and at times technical, discussion of supervenience and two-dimensional semantics, and also incorporates a variety of well-known thought experiments concerning zombies, inverted qualia, and so on. Various aspects of the argument in *The Conscious Mind* are then explored in more detail in articles collected in *The Character of Consciousness* (2010), yielding what Chalmers refers to as 'The Explanatory Argument', 'The Conceivability Argument', 'The Knowledge Argument', and the 'Two-Dimensional Argument'. Further formulations can be found in a number of discussions in which Chalmers has participated, such as those relating to his Mind and Modality class (1999) and his exchanges with Carruthers and Schier (2012a). This multiplicity of arguments raises the problem of just where to concentrate in a critique of Chalmers' case against materialism.

Thirdly, there is the contentious issue as to whether or not Chalmers begs the question against materialism. For example, it might be claimed that his argument relies on an implicit assumption that phenomenal properties are intrinsically non-physical properties, or that his distinction between the hard and easy problems takes for granted a dualistic conception of the world from the outset.[3] Now

[3] For two recent suggestions that Chalmers relies on a question-begging intuition that physical or functional accounts of experience must always leave out the phenomenal aspects, see Carruthers & Schier (2014) and Henderson (2014).

Chalmers is certainly not to be counted among those who claim that it is just *obvious* that experience is not the sort of thing that can be explained in terms of the physical world. In the introduction to *The Conscious Mind* (1996, p. xiv), he wrote that, in terms of temperament, he is 'strongly inclined toward materialist reductive explanation', and therefore his conclusions – that such an explanation cannot possibly succeed in the case of consciousness, and that we must therefore consider some form of dualism – are *conclusions* in the strongest sense, in that they were reached, reluctantly, as a result of the considerations which he sets forth in his book. Given this, we would *expect* Chalmers to set out an argument which avoids any assumption which begs the question against materialism; and I will argue in the next section that this is exactly what he does.

A fourth problem for a critique of Chalmers' argument(s) is that much of what he has to say about consciousness is written on the understanding that he has indeed established that the physical must be irreducible to the phenomenal. A case in point is his lengthy discussion of the nature of phenomenal concepts (2010, chapter 8), where he states at the outset that his discussion is premised upon 'the view that there are phenomenal properties ... that type mental states by what it is like to have them, and that phenomenal properties are not conceptually reducible to physical or functional properties'. It is therefore important in any critique to distinguish between (i) views that Chalmers holds, or comments that he makes, *as a result of* reaching the conclusion that materialism must be false, and (ii) views or comments which explicitly or implicitly act as *premises* for his argument(s) to that conclusion.

A final obstacle is that of disentangling two different aspects of conscious experience: the *qualitative* aspect of an experience, i.e. *what it is like*, and the *subjective* aspect, i.e. the fact that it is like something *for a particular organism* (Levine, 2001, pp. 6-7; Kriegel, 2009). For example, when I observe the sky on a clear day my experience has a *bluish* aspect, but it is also blue *for me*. For the moment, I simply want to note the distinction between these two different aspects of an experience – I shall have more to say about it later on.

Given these difficulties, what follows might be thought of as an *archaeology* of Chalmers' case against materialism – an investigation which aims to uncover the underlying form of his arguments, together with their presuppositions and linkages. Now some might be dubious as to the value of submitting Chalmers' arguments to detailed scrutiny: indeed, there are many philosophers who consider those arguments to be misleading, or even valueless. For example, Darren Hutchinson (2013) accuses Chalmers of 'wasting [his] time, as well as the time of others through perpetuating pointless dialectical debates', and Daniel Dennett has complained, with Chalmers evidently in mind, of a resurgence in armchair philosophy of mind that is 'not worth a damn'.[4] However, if the last 20 years

[4] Dennett's tirade, from the *Moving Naturalism Forward* conference in 2012, is worth quoting in full. Referring back to an earlier comment about people who continue to 'moan about the Hard Problem', Dennett says: 'I am just appalled to see how, in spite of what I think *is* the progress we've made in the last 25 years, there is this sort of retrograde gang, including some young ones, that are going back to old-fashioned armchair philosophy of mind with relish and eagerness. It's just sickening,

is anything to go by, a strategy of benign neglect (or even outright contempt) towards Chalmers' arguments is unlikely to be productive in terms of advancing the materialist cause. Hence the approach that I adopt here, which is to bite the bullet and get to grips with the arguments as Chalmers actually sets them out.

2: The form of the argument

Chalmers' case against materialism derives from the way in which he understands the nature of explanation: namely, that explanation depends upon the extent to which links can be established between concepts such that one set of facts entails another set of facts. What would therefore be required for a materialist account of experience is 'some sort of conceptual hook' (2010, p. 123) between our phenomenal concepts and physical or functional concepts; and it is precisely this 'conceptual hook', claims Chalmers, that is missing. In this section I will describe the argument as Chalmers sets it out, firstly in 'Facing Up' and secondly in *The Conscious Mind*, and analyse its underlying form. In the following section, I will explain how that argument fails to establish the impossibility of materialism because it does not adequately take into account a key aspect of Chalmers' own theory of meaning – that what matters for a concept's inferential role is not some description or definition, but the concept's *intension*. Since my aim in this section and the next is to demonstrate that Chalmers' argument is not successful *even if taken on its own terms*, I will be accepting (but only for the sake of argument, and only in sections 2 and 3):

> (i) Chalmers' account (1996, p. 41) of what constitutes materialism – namely, that materialism is true if all facts, including phenomenal facts, *logically supervene* on the physical facts;
>
> (ii) His claim (1996, pp. 36, 47-48) that a materialist explanation of experience must therefore be a *reductive* explanation, requiring some form of conceptual link between phenomenal facts and physical facts, such that there is an *a priori entailment* from the latter to the former; and
>
> (iii) His assertion that the only basis for such a reductive explanation that 'seems even remotely tenable' (1996, p. 104) is the same sort which works, in principle, in the cognitive sciences, i.e. some form of *functional* analysis.

Now Chalmers has characterized the proponents of materialism as belonging to two different categories. 'Type-A' materialists (e.g. Dennett, 1991; Churchland, 1996) reject the notion of a 'hard problem' because they hold that the supposed conceptual or epistemological gap between the phenomenal and the physical is illusory. Others, 'type-B' materialists (e.g. Block and Stalnaker, 1999;

---

because their work isn't worth anything, and they lure in other people to do it. It's cute, it's clever, and it's not worth a damn' (taken from a video of the conference's first session, available at https://www.youtube.com/watch?v=Ju4C_ITlBsU).

Loar, 1997; Papineau, 2002; Perry, 2001), accept that there is an epistemological gap, but maintain that this does not imply the existence of an ontological gap and the falsity of materialism. It should be clear that because I am accepting, at least for the time being, Chalmers' claim that materialism requires a priori entailment from the physical or functional facts to the phenomenal facts, my critique of his argument in this and the following section will be from the perspective of a *type-A* materialist. In section 4, however, the discussion will switch from a type-A perspective to a type-B perspective.

(a) The argument in 'Facing Up'

Chalmers begins with his distinction between the easy problems of consciousness – such as discrimination or reportability – which can be resolved by the normal techniques of cognitive science, and the hard problem, which relates to conscious *experience* and resists this type of explanation. We can characterize experience in terms of *what it is like* to feel a pain, see the colour red, and so on, and it is surely this, suggests Chalmers (2010, p. 5), which is the important problem of consciousness. The easy problems are easy, he explains (2010, pp. 6-8), because each of the relevant phenomena can be characterized in terms of the performance of a *function*, for which we just need to specify some physical mechanism or other, whereas the phenomenon of experience has a completely different nature, one which appears to have nothing at all to do with the performance of any function. At this stage, it is all too easy for the reader to assume that the distinction between the easy problems and the hard problem is simply setting out the problematic. But in fact Chalmers treats the distinction as providing a premise for his argument in 'Facing Up,' a premise which I will parse thus: experience is the one mental phenomenon which is not explainable *as* a function.

The difference between the two types of problem can be understood in terms of the contrasting nature of our concepts of the relevant phenomena (2010, pp. 7-8). Our understanding of what, say, 'discrimination' or 'reportability' actually *mean* can be stated in functional terms, and therefore we have a good reason to expect the easy problems to be solvable by cognitive science. But a functional account of conscious experience simply seems to fail: even after we have explained all the mental functions in the vicinity there is still an additional question about why there is also an experience of some sort (2010, p. 8). We therefore have a gap between our functional descriptions and our phenomenal descriptions (2010, p. 8) – an *explanatory gap*, to use Levine's phrase – and this is evidenced by the fact that it is conceptually coherent that our brains could carry out all the physical and functional processes we like, and yet there might be no conscious experience (2010, p. 14).

To cross such a gap, says Chalmers, we require more resources than can be discovered on the physical side alone: some 'extra ingredient' (2010, p. 13) must be found. Physical accounts in terms of structure and dynamics can explain functions by specifying a physical mechanism, but otherwise they can only yield more structure and dynamics (2010, p. 15), and such accounts therefore remain marooned on the far side of the explanatory gap.

We can therefore summarize the argument in 'Facing Up' as follows:

(1) Physical accounts can only explain functions or yet more physical structure and dynamics;

(2) But experience is not explainable *as* a function (nor, presumably, is it explainable as physical structure or dynamics); hence

(3) There can be no physical explanation of experience.

(b) The argument in 'The Conscious Mind'

Chalmers describes his book as providing a fuller and more detailed case against materialism than that contained in 'Facing Up'. The argument starts by defining materialism as the doctrine that the physical facts determine all positive facts about the world: i.e. once you have fixed the physical facts, you have fixed *all* of the facts (1996, p. 41). Chalmers then expresses this notion more formally in terms of *supervenience*: materialism is true if all facts, including phenomenal facts, *logically supervene* on the physical facts (1996, p. 41).

Now on Chalmers' understanding (1996, pp. 36, 47-48), the metaphysical notion of the logical supervenience of *facts* and the epistemological notion of reductive *explanation* are closely related: if a phenomenon is to be reductively explainable, then the property which instantiates that phenomenon must be logically supervenient on lower-level properties. If this were not the case, there would still be a question as to *why* the relevant phenomenon accompanied the lower-level properties; hence, we would not have a full explanation of that phenomenon. This view of what constitutes a reductive explanation has been criticised,[5] but I will be taking it for granted here. What it implies, says Chalmers, is that a wholly materialist explanation of experience must be a *reductive* explanation, requiring some form of conceptual link between phenomenal facts and physical facts such that there is an *a priori entailment* from the latter to the former (1996, pp. 47-48).[6]

Chalmers then suggests (1996, p. 70) three different ways in which we might establish that B-properties are logically supervenient on A-properties: (i) argue that it is not conceivable that the A-properties can be instantiated without the B-properties being instantiated (arguments from *conceivability*); (ii) argue that possession of the A-facts will enable us to know the B-facts (arguments from *epistemology*); and (iii) analyse our concept of the B-property to show that statements concerning B-facts follow logically from A-facts (arguments from *analysis*). The reasoning here seems clear enough: if B is indeed *logically* supervenient on A, then I should be able to see that conceiving of A being the case but B not being the case leads to a logical contradiction; I should also

[5] For example, by Block and Stalnaker (1999); for responses, see Chalmers & Jackson (2001) and Carruthers (2004).

[6] A reductive *explanation* of some phenomenon does not, on Chalmers understanding, necessarily imply an actual reduction of that phenomenon – in the sense that we would be able to *identify* the phenomenon with some specific lower-level phenomenon (1996, p. 43).

be able to deduce that if A is the case, then B is also the case; and finally, I should be able to provide some form of analysis of B in terms of A.

Chalmers goes on (1996, pp. 94-106) to give five specific arguments intended to show that conscious experience cannot be logically supervenient on the physical, and that any reductive explanation of experience must therefore fail. These are:

> 1. The logical possibility of zombies – where a 'zombie' is defined as a being that is physically and functionally identical to a human being, but has no conscious experience; (an argument based on conceivability).
>
> 2. The logical possibility of an inverted spectrum, e.g. that I might have had the experience of red in exactly the same physical situation in which I actually experience blue; (another argument based on conceivability).
>
> 3. Epistemic asymmetry – our knowledge about phenomenal experience is obtained mainly from our ourselves, rather than from the external world or other people; (an argument based on epistemology).
>
> 4. Frank Jackson's argument concerning 'Mary', the cognitive psychologist who has never seen any colours and is therefore supposed to lack knowledge about certain phenomenal facts, no matter how extensive her knowledge of the physical and functional facts; (another argument based on epistemology).
>
> 5. The impossibility of analyzing phenomenal experience in terms of anything else; (an argument based on analysis).

Finally, having examined the epistemological question as to whether there can be a reductive explanation of conscious experience, Chalmers reaches his ontological conclusion: if experience is not logically supervenient on the physical, then the facts about experience are *additional* facts about the world that are not determined solely by the physical facts. Hence, materialism – 'the doctrine that the physical facts about the world exhaust all the facts, in that every positive fact is entailed by the physical facts' – must be false (1996, pp. 123-124).

(c) A 'master version' of the argument

Chalmers has said (2010, p. xv) that the argument in 'Facing Up' is more fundamental to the case against materialism than the well-known thought experiments in *The Conscious Mind*, since the latter turn on points concerning structure and function that are similar to those made in 'Facing Up'. The substantive difference between these two versions of the anti-materialist argument is that in the book-length treatment Chalmers formalizes the notion of 'explanation' in terms of the a priori entailment of facts. We can therefore envisage a 'master version' of the argument against materialism which would succinctly and clearly set out the form of the argument based upon structure and function, whilst also

taking into account Chalmers' formalization of the nature of explanation. Fortunately, Chalmers himself has provided such a formulation:

> (1) Physical concepts are all structural-dispositional concepts;
>
> (2) If B truths are to be entailed a priori by structural-dispositional truths, there must be some analysis of B concepts in structural-dispositional terms;
>
> (3) There is no analysis of phenomenal concepts in structural-dispositional terms; so
>
> (4) Phenomenal truths are not entailed a priori by physical truths (Chalmers 1999)

(where 'B' stands for some domain that is to be subject to reductive explanation).

Before evaluating this 'master version' of the argument, I need to clarify one point. Chalmers varies in how he characterizes the nature of physical properties or concepts. Sometimes he describes them as being expressed in terms of 'structure and dynamics', but on other occasions he refers instead to 'structure and function', or to 'structural-dispositional' properties or concepts. In practice, little seems to depend upon the differences between these various formulations. However, it should be noted that by 'structure' Chalmers is referring specifically to *spatio-temporal* structure (2010, pp. 105-106), and not, for example, to the structure of the relations that the qualitative aspects of experiences have with respect to each other, a structure which can be conceptualized in terms of a 'quality space' (Clark, 1993).

Turning now to the three premises of the 'master version' of Chalmers' argument against materialism: to counter that argument by denying the first premise, that physical concepts are structural-dispositional in nature, would seem an unpromising choice; and the second premise is strongly implied by the nature of a priori entailment. Attention is therefore fixed on the third premise: 'There is no analysis of phenomenal concepts in structural-dispositional terms'. It is difficult to envisage an analysis of phenomenal properties or concepts that is *directly* expressed in terms of spatio-temporal structure or dynamics, and Chalmers therefore believes that the only basis for a reductive explanation of experience that is 'even remotely tenable' is the same sort which works, in principle, in the cognitive sciences, i.e. some form of *functional* analysis (1996, p. 104). As I explained earlier, I will be taking Chalmers' position here for granted, and we can therefore simplify the third premise to: *There is no analysis of phenomenal concepts in functional terms*.

Now this is a very general premise, which gives Chalmers' argument a wide scope. It does not matter what form a functional explanation takes, such as analytic functionalism or teleofunctionalism; nor does it matter whether experience is identified with higher-level functional properties or lower-level functional properties; and, finally, it does not matter how broad our definition of 'function' may be, e.g. simply an identification of states by what they do rather than by what they are (Dennett 2005, p. 17). Without *some* form of analysis of phenomenal concepts in functional terms, there would be no

possibility of any a priori entailment from functional facts to phenomenal facts, and therefore no reductive explanation of experience**.**

However, rather than mounting any argument against the possibility of functional analyses of phenomenal concepts, Chalmers simply dismisses the idea: 'To analyze consciousness in terms of some functional notion is either to change the subject or to define away the problem. One might as well define "world peace" as "a ham sandwich"' (1996, p. 105). For Chalmers, it seems, the point is as self-evident as the distinction between the easy problems and the hard problem. Now to deny that there is some mental phenomenon – experience – which needs to be explained, in addition to such phenomena as awareness or discrimination, might indeed be to 'deny the manifest' (2010, p. 112); but should why an *analysis* of experience in functional terms be equally implausible?

The reason is that Chalmers is considering every aspect of an experience *as an explanandum in its own right* – as a phenomenon that is in need of explanation. For example, my experience when I look at the sky on a clear day has a bluish aspect to it, and even if I associate that aspect of my experience with cognitive abilities such as, say, recognition and discrimination, or with the relational properties of hue, saturation and lightness which blue has with respect to other colours, it is nevertheless possible to 'carve out', as Peter Carruthers has put it (2004, p. 163), the purely phenomenal aspect, the aspect which is characterized in terms of *what it is like*, to form a separate concept that can *only* be characterized phenomenally – as, say, 'bluish'. It would not appear to be possible for *that* concept to have anything of the functional about it which could provide the basis for a reductive explanation. The orthogonal nature of the two different sorts of concept would seem to preclude any 'conceptual hook' (2010, p. 123) by which to link my concept of 'bluish' to functional concepts, and some such linkage is a prerequisite for a priori entailment.

To accept the distinction between the easy problems and the hard problem is therefore *ipso facto* to accept that there is indeed some aspect of an experience which, as an explanandum, is characterized phenomenally and *only* phenomenally. To go on to claim that one can nevertheless provide a functional analysis of our concept of that aspect of an experience is, on the face of it, simply incoherent – which explains why Chalmers says that any such purported analysis must either define the problem away or change the subject matter. Any proposed explanation of experiences in physical or functional terms – whether by way of emergence, causality, a posteriori identity, constitution, or even co-constitution – therefore raises a further question: *why* should some specific phenomenal aspect of an experience be emergent from, caused by, a posteriori identical with, constituted by, or co-constituted with, any particular set of physical or functional properties? To this question, it seems, the materialist can have no possibility of an answer.

(d) The strength of the argument

We can now see the strength of Chalmers' argument, in that it relies only on the nature of phenomenal and functional concepts, the nature of explanation, and an assumption which could only be denied by denying the manifest, namely that there is some phenomenal aspect to any experience. It therefore avoids any contestable premise concerning the nature or ontological status of experience or of phenomenal properties – a premise which the materialist could simply claim begs the question so far as they are concerned. This understanding of the form of Chalmers' argument is consistent with the account which Chalmers himself provided during a discussion at the 2012 Consciousness Online conference. In that discussion, he explained that:

> [t]he claim that there is [a hard problem distinct from the easy problems] does *not* rule out many nearby physicalist views, for example ... the view that consciousness can be explained in terms of structure and function. Someone might say: I agree that the problem of explaining conscious experience is distinct from the problem of explaining the various structures and functions, but I think one can explain it in terms of structure and function all the same. I think those people are wrong, but they aren't contradicting themselves. (2012a)

Hence the claim that there is a hard problem is simply the claim that there is some mental phenomenon, in addition to functional phenomena, that needs an explanation. The problem is 'hard' simply because the phenomenon in question is not characterized or identified at the outset in functional or structural terms, and this of itself does not rule out a functional or structural explanation.

In his comments at Consciousness Online, Chalmers goes on to say that what *does* rule out a position which accepts the existence of the hard problem, but maintains that experience may nevertheless be explainable in terms of structure and function, is the combination of the following: (i) that experience is a problem distinct from the functionally characterized easy problems, *and* (ii) that 'structure and function explain only structure and function' (2012a). Now taken as it stands, (ii) might seem to beg the question against a functional or structural explanation of experience. But as I have described, it is justified so far as Chalmers is concerned because the differing nature of the relevant concepts prevents any a priori entailment of the phenomenal facts by the functional or structural facts. The key to Chalmers' case against materialism is therefore the orthogonal nature of phenomenal and functional concepts:

> ... if I were to rest on any point as supporting the central burden, it might be ... the conceptual distance (at least in a priori space) between functional and phenomenal concepts. (1999)

Now that we have correctly understood the form of the argument, we should be able to see more clearly whether it succeeds in establishing its conclusion – and I shall argue in the next section that it does not.

3: Why the 'master version' of the argument fails

(a) Epistemic intensions and explanation

We can see from the preceding discussion that Chalmers' argument against the possibility of materialism rests on three claims: firstly, that there is some mental phenomenon in need of explanation in addition to functionally characterized phenomena (it is the rejection of *this* claim that Chalmers describes as the denial of 'the manifest'); secondly, that a materialist explanation of that phenomenon requires that the physical or functional facts a priori entail the phenomenal facts; and thirdly, that the difference between functional concepts and phenomenal concepts prevents any such a priori entailment.

However, this last claim suffers from one major difficulty: on Chalmers' own account of meaning (e.g. 1996, pp. 54), what is important for a concept's inferential role is the concept's intension, a function which specifies how that concept applies in different situations. Hence, the fact that we can characterize an aspect of an experience in a purely phenomenal manner, for example as the 'bluish' aspect of my experience when I look at the sky, is not of itself what is important so far as a priori entailment is concerned. What Chalmers actually needs to demonstrate, if he is to prove that a phenomenal concept cannot have an inferential role which could conceivably ground a priori entailments from functional facts, is that the concept's *intension* cannot be formulated or described in functional terms.[7] I shall therefore now consider the nature of intensions as Chalmers understands them, and how he deals with the intensions of phenomenal concepts, before concluding as to the implications for his argument against materialism.

Intensions are functions from possible situations, usually defined in terms of possible worlds, to extensions of an expression; they can apply to concepts or thoughts, as well as the utterance of a term or expression (2002a, pp. 176-177). Chalmers utilizes intensions within a two-dimensional framework in which expressions have two main kinds of meaning – an epistemic (or primary) intension and a subjunctive (or secondary) intension (2002a; 2006; 2010, Appendix).[8] An *epistemic* intension is the function that we obtain by evaluating an expression in every logically possible world, where we consider those worlds as actual: if the world *were* that way, what would the relevant expression pick out? The usual example (e.g. 1996, pp. 57-60) is 'water': it is possible, *given all that we might know a priori*, that water might have been XYZ rather than H_2O. Therefore the epistemic intension of 'water' would pick out H_2O in our world and XYZ in a world where the clear, drinkable liquid in oceans and rivers is made up of XYZ. The other type of intension is a *subjunctive* intension, which is the result of

[7] This can be taken as either denying premise (2) in Chalmers' four-point argument quoted in section 2(c) or denying premise (3), depending on whether one considers a description of an intension of a concept to constitute an analysis of that concept.

[8] Chalmers (2002a; 2006) has refined the account that he provided in *The Conscious Mind* (1996), where he used the terms *primary intension* and *secondary intension*, and now usually refers instead to *epistemic intension* and *subjunctive intension*. For my purposes here, there is no significant difference.

evaluating every possible world as counterfactual: given the way the world actually is, what *would* the expression have picked out had the world been some other way? So given that water is actually H_2O, the subjunctive intension of ‘water’ picks out H_2O in all possible worlds.

Chalmers uses his two-dimensional framework (1996, pp. 131-140; 2010, Chapter 6) in order to counter the arguments of type-B materialists, who claim that epistemological arguments cannot entail the metaphysical conclusion that materialism is false. For example, he analyzes the divergence of conceivability and identity in the case of natural kind concepts such as ‘water’, and argues that this sort of divergence cannot occur in the case of phenomenal concepts. For a natural kind term, the epistemic and subjunctive intensions can pick out different extensions: water, therefore, is necessarily H_2O (by the subjunctive intension), even if it is conceivable a priori that it might not have been H_2O (by the epistemic intension). But in the case of phenomenal terms, there would seem to be no possibility of any such divergence, since what it takes for a phenomenal concept to apply in any possible world considered as actual is for an experience to have a particular feel, and what it takes for that phenomenal concept to apply in a counterfactual world is also for an experience to have that particular feel (1996, p. 133).

However, I am for the time being forgoing any objection to Chalmers’ argument from a type-B perspective: as noted earlier, I am taking for granted (at least for now) Chalmers’ claims concerning the relationships between materialism, logical supervenience, and a priori entailment. Hence my interest is not with two-dimensional semantics *per se*, but with how a concept’s intension governs its role in inference – and specifically its role in a priori entailment. Chalmers notes that the epistemic intension is grounded in epistemic necessitation, i.e. a priority, and the subjunctive intension in metaphysical necessitation (2010, p. 550); for my purposes, therefore, it is the *epistemic* intension which is relevant.

On Chalmers’ understanding, a term or expression is associated with certain criteria, very possibly implicit, which reflect the way in which a subject applies that term or expression to the world. There are very many different ways in which the world might have turned out to be, providing us with an *epistemic space* of possibilities to consider, but there is also – claims Chalmers – *scrutability* of truth and reference, i.e. given sufficient information about a possible world, a subject can make rational judgments as to what a particular term applies to in that world, and whether their utterances involving that term are true (2006, pp. 75-93). It is epistemic space and scrutability which suggest to Chalmers (2006, p. 75) the idea of an epistemic intension operating as a function from epistemic possibilities to extensions. For example, the epistemic intension of ‘water’ can be approximately described[9] as ‘the dominant clear, drinkable liquid in the environment’ (1996, p. 57), and we can arrive at this

[9] Although an intension is not itself identical to a description or definition, Chalmers nevertheless views them (2010, p. 556) as being describable in *approximate* terms, even though such descriptions will often be imperfect.

formulation by considering the epistemic space of possibilities and determining what the term ‘water’ applies to in a variety of possible worlds:

> *The true intension can be determined only from detailed consideration of specific scenarios*: What would we say if the world turned out this way? What would we say if it turned out that way? For example, if it had turned out that the liquid in lakes was H_2O and the liquid in oceans was XYZ, then we probably would have said that both were water; if the stuff in oceans and lakes was a mixture of 95 percent A and 5 percent B, we would probably have said that A but not B was water; if it turned out that a substance neither clear nor drinkable bore an appropriate microphysical relation to the clear, drinkable liquid in our environment, we would probably call that substance ‘water’ too (as we do in the case of ice or of ‘dirty water’). (1996, pp. 57-58, emphasis added)

Because an epistemic intension is not a description or a definition (2002a, pp. 148-149) but a function that is derived by considering how a subject would apply the relevant expression in specific situations, it is ultimately grounded in the expression’s inferential role:

> [F]or any expression we use, then given sufficient information about the actual world, certain judgments using the expression will be irrational, and certain other judgments using the expression will be rational. ... It is this sort of inferential role that grounds the [epistemic] intension of an arbitrary expression (as used by an arbitrary speaker). A given sentence token will be associated with a raft of conditional rational judgments across a wide variety of scenarios. This raft of conditional judgments corresponds to the sentence's [epistemic] intension. (2010, p. 555)

This provides us with a way of understanding how epistemic intensions relate to reductive explanation: where we have a reductive explanation of some phenomenon or other, this is possible because the epistemic intension of our term for that phenomenon, and hence that term’s role in inference, is capable of being described by means of *other* terms. The expression ‘water’, for example, is not itself characterizable in dispositional terms; but we can investigate its epistemic intension and describe that intension using expressions such as ‘clear’, ‘drinkable’, ‘liquid’, etc., which *are* characterizable in dispositional terms. Those expressions can then be characterized, either directly or by way of expressions which describe *their* epistemic intensions, in terms of causal roles. It is then an a posteriori matter of scientific discovery as to what it is that actually fulfills those causal roles; in our world it is H_2O molecules, and once we know that fact we can obtain a reductive explanation of the behaviour of water in terms of H_2O molecules.

Scrutability of reference implies that, even if the criteria for applying a particular term are implicit and discoverable only by considering our dispositions to apply that term over a range of epistemic possibilities, there nevertheless must *be* such criteria. We can therefore view such applicability criteria as grounding the type of analysis that would be required for a reductive account of some phenomenon or other. So if it turns out to be the case that the applicability criteria for phenomenal concepts are in

fact expressible in functional terms, then an a priori entailment from the functional facts to phenomenal facts would be a possibility after all.[10]

One implication is that the *form* which we consider a phenomenal concept to have (that it is a recognitional concept, for example, or an indexical concept) is not directly significant for its role in a priori entailment. Consider, for example, Papineau's formulation (2007, p. 112) '*the experience:* —, in which the gap is filled either by a current experience or by an imaginative recreation of an experience'. This envisages phenomenal concepts as a form of quotational concept, whereby a token of the phenomenal quality actually enters into the concept.[11] It might be suspected that the inclusion of the phenomenal token would prevent the concept from playing any role in logical inference to or from facts that are expressed using non-phenomenal concepts. However, what is relevant for a priori entailment, on Chalmers' understanding, are the implicit criteria by which we *apply* the concept, rather than the fact that the concept itself incorporates a token of an experience; hence, there is no reason in principle why a quotational concept such as Papineau's should not be capable of sustaining a priori entailments from functional facts to phenomenal facts.

(b) Chalmers' treatment of the epistemic intensions of phenomenal concepts

The notion of an epistemic intension is clearly central to Chalmers' understanding of the role that concepts play in our cognitive activities. Yet his investigation of the applicability criteria for phenomenal concepts is remarkably limited. We might have expected him to follow his own injunction that an intension is to be ascertained by means of a detailed consideration of specific scenarios – as in the example he gives of 'water'. Particularly relevant, one would have thought, would have been scenarios in which those functionally characterized mental phenomena that normally accompany our experience of a phenomenal quality are not present – just as, in the case of water, Chalmers considers possible worlds with oceans made of clear liquids that are not drinkable, worlds with clear, drinkable liquids constituted by XYZ, and so on.

In fact, when Chalmers does give a detailed account of phenomenal concepts (2010, chapter 8), his discussion explicitly *assumes* at the outset that phenomenal properties are not capable of being conceptually reduced to physical or functional properties. This is an assumption that goes beyond the identification of experience as a phenomenon which is characterized in terms of 'what it is like':

> The discussion that follows is premised upon what I call 'phenomenal realism': the view that there are phenomenal properties (or phenomenal qualities, or qualia) – properties that type mental states by what it is like to have them – *and* that phenomenal properties are not conceptually reducible to physical or functional properties ... (2010, p. 252, emphasis added)

[10] At this point it might be objected that the applicability criterion for my concept of, say, the bluish aspect of my experience is surely just that it seems blue to me. But this is too quick: someone might equally claim that the applicability criterion for 'water' is 'watery stuff'. We can go on to ask what the applicability criteria are for 'seems blue to me', just as we could for 'watery stuff'.

[11] Similar accounts of phenomenal concepts have been given by, among others, Block (2006) and Chalmers (2003).

Chalmers then distinguishes between various types of concept that we have in respect of a particular phenomenal quality (2010, pp. 254-260). Taking the experience of red as an example, there are concepts which have their referents fixed by relation to external objects, such as the *community relational concept* (which can be glossed as 'the phenomenal quality which paradigmatically red objects typically cause in my community's normal subjects') and the *individual relational concept* (glossed as 'the phenomenal quality which paradigmatically red objects typically cause in me'); but there is also a *pure phenomenal concept*, which does not pick out its reference relationally but does so directly in terms of red's 'intrinsic phenomenal nature' (2010, p. 256).[12]

However, what Chalmers does not do in this discussion of phenomenal concepts, or elsewhere, is critically investigate the applicability criteria for our phenomenal concepts by considering a range of possible scenarios. This omission has led John Perry (2004; see also Alward, 2004) to claim that Chalmers takes the epistemic intensions of phenomenal concepts to be the phenomenal properties themselves; for example, that the intension of 'pain' is pain itself. But Perry is not quite right: it is not that Chalmers takes pain to be the intension of itself, but rather that he takes the applicability criterion for our concept of the experience of pain to be simply the instantiation, in someone's experience, of the phenomenal property of pain:

> It's true that I think the primary intension of pure phenomenal concepts pick out the corresponding phenomenal qualities in all worlds, so that the primary intension of [the] pure phenomenal concept of 'pain' picks out the phenomenal quality of pain in all worlds. (I'd prefer to put it this way than to say that the primary intension of 'pain' is pain, which is the way Perry puts it. I don't think pain is an intension, for a start.) (2004)

In fact Chalmers' position here is not entirely unexpected, given the way in which he employs two-dimensional semantics in his critique of materialism. As I described above, one of his principal uses of two-dimensionalism (e.g. 1996, pp. 132-140; 2010, Chapter 6) is to counter the views of type-B materialists, who accept that there is a conceptual, or epistemic, gap between the phenomenal and the physical but deny that this implies a metaphysical gap. Both parties to this particular debate accept the existence of an epistemic gap, and at this point in the dialectic Chalmers has already established to his own satisfaction the conceptual independence of phenomenal properties and hence their non-

---

[12] Chalmers does not define precisely how he uses – either here or elsewhere – the term 'intrinsic'. However, he does tend to contrast 'intrinsic' with 'relational', as in the discussion quoted above, and I therefore take the claim that phenomenal qualities have 'intrinsic phenomenal natures' to be equivalent to the claim that phenomenal properties are intrinsic properties. There is some debate about just what it is that makes a property 'intrinsic' as opposed to 'extrinsic', although the basic notion is well-described by Yablo (1999): 'You know what an intrinsic property is: it's a property that a thing has (or lacks) regardless of what may be going on outside of itself'. One will frequently find 'intrinsic property' equated with 'non-relational property', but this usage has been questioned (see, for example, the discussion in Weatherson & Marshall, 2012). However, nothing in what follows will depend upon a precise definition of 'intrinsic'. If, on the other hand, all that Chalmers means by 'intrinsic phenomenal nature', is something like 'the nature of the phenomenal property, whatever it may be', then this does not exclude the possibility that phenomenal properties might be relational properties, or that the applicability criteria for our concepts of those properties might by describable in physical or functional terms. In any event, the critical question for Chalmers' argument against materialism is: 'what *are* the applicability criteria for phenomenal concepts?'

reducibility to physical or functional properties. It would therefore be quite natural for him to take it for granted, in any discussion of type-B materialism, that phenomenal properties instantiate intrinsic phenomenal properties and that such instantiation acts as the applicability criterion for the relevant concept.

But although Chalmers' treatment of the intensions of phenomenal concepts is not surprising, it is unsatisfactory in two respects. Firstly: as I noted above, it ignores his own statements about how we should go about evaluating epistemic intensions more generally, and therefore begs the question as to what might be revealed by an investigation into how phenomenal concepts are actually applied in a variety of scenarios. Secondly: as we saw in section 2, the form taken by the argument against materialism in 'Facing Up' and *The Conscious Mind* proceeds from a straightforward identification of conscious experience as a phenomenon that is in need of explanation, distinct from other mental phenomena, and therefore provides no warrant for Chalmers to assume the existence of intrinsic phenomenal properties *prior to having established that argument's conclusion*. Indeed, Chalmers himself suggests that his case against materialism does not rely upon any premise, whether explicit or implicit, as to whether phenomenal properties are intrinsic properties. Consider, for example, his definition of 'qualia':

> In my usage, qualia are simply those properties that characterize conscious states according to *what it is like* to have them. *The definition does not build in any further substantive requirements such as the requirement that qualia are intrinsic* or nonintentional. If qualia are intrinsic or nonintentional, this will be a substantive rather than a definitional point (so the claim that the properties of consciousness are non-intrinsic or that they are wholly intentional should not be taken to entail that there are no qualia). (2010, pp. 104-105, emphasis added)[13]

Now Chalmers' argument against materialism is also framed around the claim that there are mental phenomena which are identified in terms of 'what it is like' to have them: it is precisely these phenomena that constitute the 'hard problem'. So neither Chalmers' characterization of qualia, nor his characterization of the hard problem, go any further than the claim that there are mental phenomena which are identified in terms of 'what it is like'; neither characterization *ipso facto* entails the claim that phenomenal properties are *intrinsic* properties.

Similarly, when Chalmers uses the principle that 'from structure and dynamics, one can infer only structure and dynamics' in order to draw the conclusion that phenomenal facts cannot be inferred from physical facts, he insists that 'claims about intrinsic properties *play no role* in the argument I have given' (2010, p. 121, emphasis added). Accepting Chalmers' argument against materialism

---

[13] Chalmers' definition of qualia in his book *The Conscious Mind* (1996, p. 359) also excludes any substantive, additional requirement: 'Different authors use the term "qualia" in different ways. I use the term in what I think is the standard way, to refer to those properties of mental states that type those states by what it is like to have them. In using the term, I do not mean to make any immediate commitment on further issues, such as whether qualia are incorrigibly knowable, whether they are intentional properties, and so on.'

might very well lead one to *conclude* that phenomenal qualities are indeed intrinsic properties, but their intrinsicality does not feature, and is not required, as a premise in that argument: the claim that phenomenal concepts are not analyzable in functional terms suffices.

Of course, it is clearly possible to construct a case against materialism around the claim that phenomenal properties are intrinsic properties and that the applicability criteria for our phenomenal concepts are simply the instantiation of such properties. Such an argument would depend upon some form of premise concerning the nature of phenomenal properties: either an ontological premise (that experiences *have* intrinsic phenomenal properties) or an epistemological premise (that we have indubitable *knowledge* about intrinsic phenomenal properties). Hence, whatever its other merits or demerits, this type of argument would be open to the objection that it proceeds from a characterization of experience which is inherently dualistic and which therefore begs the question against the materialist. Chalmers himself, however, makes no such mistake. As I noted at the end of section 2, the strength of his argument is precisely that it does *not* depend upon some assumption concerning the nature of phenomenal properties, but upon the claim that our experiences present phenomenally qualitative aspects as explananda – a claim which cannot be questioned without appearing to deny the manifest. But in the case of a premise concerning intrinsic phenomenal properties, it would be open to a materialist to argue that they can contest *that* premise *without* denying the manifest.

In any event, the problem for the ‘master version’ of Chalmers’ argument against materialism is this: why can’t the implicit applicability criteria for a phenomenal concept be described in terms that relate to cognitive functions? Chalmers gives us no reason to think that this is not a possibility; and if it *is* a possibility, then an a priori entailment from functional facts to phenomenal facts cannot be ruled out. For example, it might be that an investigation into how someone would apply their concept of *sky blue* to their experiences would reveal that they would do so if they could *recognize* the colour that they are experiencing as that which they term ‘sky blue’ and if they could *discriminate its relational properties* with respect to other colours as being those which they expect ‘sky blue’ to have, e.g. that there is nothing red or green about it, that it is not as deeply saturated a blue as navy blue, and that it is quite light in the same way that yellow is. If their experience meets those criteria, then they would say that their concept of *sky blue* applies to their experience.

Chalmers’ criticism of materialists such as Dennett (e.g. 2010, pp. 9-13) is that they explain certain cognitive functions and then conflate, either explicitly or implicitly, those cognitive functions with phenomenal experience. What is being suggested in this simple example, however, is entirely different. The explanandum under consideration is indeed a phenomenal aspect of experience, the aspect which we can characterize as *sky bluish*, and terms such as ‘recognize’ and ‘discriminate’ are being used within the descriptions of the implicit applicability criteria for our concept of that phenomenal aspect, rather than as explananda in their own right. (It might be thought that the applicability criteria suggested in this example are not wholly characterized non-phenomenally, since

they refer to phenomenal qualities such as sky blue, red, green and so on. But these references can be to 'the colour I term *sky blue*', 'the colour I term *red*', and so on, where the word 'colour' simply refers to an area of the subject's visual field being filled in some distinguishable manner or other.)

This suggestion as to the applicability criteria for phenomenal concepts gains plausibility when we consider what Chalmers refers to as 'dancing qualia' (1996, pp. 266-274). Suppose that there *are* in fact intrinsic phenomenal properties. Then the following sort of situation appears to be conceivable: I am looking at the sky and the intrinsic property of my experience switches from blue to red, yet I do not notice any change. I carry on observing the sky and I continue to recognize the colour as that which I term 'blue', and to discriminate its relational properties to other colours to be those which blue has, *even though my experience now has the intrinsic property of phenomenal redness*. Chalmers has referred to this type of situation as being indicative of a 'radical dissociation between consciousness and cognition', but nevertheless believes that it is logically possible – albeit 'only just logically possible' (1996, p. 269). Assuming that such a bizarre situation *is* possible, then the question which needs to be considered is this: which of my concepts would I say applied to my experience – my concept of 'bluish', or my concept of 'reddish'? Surely I would say that my concept of 'bluish' applies. In other words, my criteria for applying a phenomenal concept to my own experiences would *not* be the instantiation of the supposed intrinsic phenomenal property. This result is not entirely unexpected, once we reflect on the fact that the applicability criteria for the use of a concept is something which concerns *cognition*, rather than experience.

Of course, the anti-materialist may try and claim that the extreme nature of the dancing qualia scenario means that I should discount the result. But it is precisely unusual situations which may reveal something unexpected about our concepts, as in Chalmers' example of a world where water is made up of molecules of XYZ rather than H_2O. The example of dancing qualia shows that the need for a critical examination of the applicability criteria for phenomenal concepts cannot be preempted by a claim that there are intrinsic phenomenal properties – a claim that would in any event beg the question against a materialist explanation of experience.

(c) Scrutability and the failure of the 'master version' of the argument

Chalmers' error lies in not fully taking into account the distinction between (i) a description or characterization of a phenomenal property, and (ii) the epistemic intension of our concept of that property. The fact that we characterize or identify the phenomenal aspect of an experience without using functional terms, e.g. as 'bluish', does not imply that functional terms cannot adequately express the applicability criteria that might be revealed through a critical investigation of the relevant concept's intension.[14]

[14] On Ray Jackendoff's account (2012), meaning as such is something that we are not usually conscious of, even though it does the work of establishing inference and reference. If this is right, then it will be easy to confuse a description or

What is it then, which makes it possible to conceptually link functional concepts and phenomenal concepts – despite their apparently orthogonal nature? After all, this is the sort of difficulty which has led many materialists to adopt some version of type-B materialism: that is, to accept that there is an epistemological gap between the phenomenal and the physical or functional, and then argue that this does not imply any ontological gap. What in fact mediates the connection between the two types of concept is Chalmers' thesis of *scrutability*: given sufficient information about the world, we are able to make rational judgments about what our expressions refer to and whether statements incorporating those expressions are true. We therefore have an understanding of how to apply our concepts in different situations, and should be capable (by considering a range of epistemic possibilities) of ascertaining their implicit applicability criteria. But precisely because those criteria *are* implicit, and the qualitative aspects of our experiences so ineffable and so familiar, it may not be evident to us that functional facts and phenomenal facts can be conceptually linked via the applicability criteria for phenomenal concepts.

Of course, Chalmers' scrutability thesis is debatable. But it also plays a key role in supporting his claim that, for materialism to be true, the phenomenal facts must be a priori entailed by the physical facts. Scrutability ensures, that with *sufficient* information about the world, we can always make rational judgments, i.e. judgments that require no further empirical warrant and are therefore a priori, concerning reference and truth (Chalmers, 2010, pp. 553-556); hence, if materialism were true, the phenomenal facts should be (at least in principle) deducible a priori from *some* set of physical facts.[15] If, contra Chalmers, the scrutability thesis is rejected, then a lack of a priori entailment from physical or functional facts to phenomenal facts would no longer entail that materialism is false. Chalmers' argument against the possibility of materialism would then fail for the reason put forward by type-B materialists – that an epistemological gap does not imply an ontological gap.

We might wonder just why it is that Chalmers believes the scrutability thesis to be true. Philip Goff points out (2011, pp. 193-196) that Chalmers' claim that concepts have epistemic intensions which can be evaluated a priori amounts to a rejection of 'radically opaque' concepts – that is, concepts which reveal neither essential properties, nor accidental properties that allow us to establish the concept's referent in the actual world. For example, our concept of water is *not* radically opaque, because it reveals to us the accidental property of being the dominant clear, drinkable liquid in our

---

characterization with a concept's meaning. It also suggests that the applicability criteria for a concept need to be carefully investigated, even if – *especially* if – the meaning appears obvious: '[t]here is nothing more deceptive than an obvious fact,' as Sherlock Holmes put it (Doyle, 1981, p. 204).

[15] It is therefore rather ironic that Chalmers' views seem to be of particular appeal to those who are concerned about reductive explanations or the hubris of believing that human beings can understand and explain anything at all. Chalmers maintains that if experience were really part of the material world, then we should be able to deduce the facts about our experiences from the physical facts; but since we can't make this sort of a priori deduction, experienced qualities cannot be part of the material world, and must instead be fundamental entities akin to the fundamental entities of physics. In short: experience must be fundamental because we humans cannot (yet) reductively explain it. What argument could be more rationalistic or hubristic than *that*?

environment, and this allows us to establish the referent of ‘water’ as H_2O in our world and XYZ in a world where the predominant clear, drinkable liquid is XYZ (Goff 2011, p. 195). But if phenomenal concepts *were* radically opaque, then their referents might indeed be physical or functional properties, *even if* there were no a priori entailments between physical or functional facts and phenomenal facts. Many philosophers *do* doubt the notion that our concepts provide us with some form of direct access to properties – David Papineau, for example, argues that the approach taken by Chalmers rests on a ‘highly suspicious’ notion of direct acquaintance:

> It assumes some mode of thought where objects become completely transparent to the mind and all their essential properties are thereby laid bare. It is hard not to see this as inspired by some misplaced visual model, in which we are able to peer in at some immaculately illuminated scene. (Papineau 2006, p. 102)

So how does Chalmers attempt to establish the truth of the scrutability thesis? His method is to consider various types of fact (e.g. macro-physical, economic, biological, sociological – even moral or aesthetic facts, if one takes a realist attitude to them) and argue that scrutability is a near-universal feature of our judgments about the way the world is. The thesis itself is not mentioned in *The Conscious Mind*, but is implicit in Chalmers’ argument (1996, pp. 71-89) that all facts (such as facts concerning life and economic prosperity) are logically supervenient on the physical facts – except for facts about conscious experience, indexical facts, and negative existential facts. He takes the same approach in a paper co-authored with Frank Jackson (2010, Chapter 7) which argues that all macrophysical truths are entailed by the sum of all microphysical truths and phenomenal truths,[16] plus a ‘locating’ statement and a ‘that’s all’ statement. This is exactly what one would expect if the scrutability thesis were true – sufficient information about the world will always enable us to determine the referents of expressions and the truth and falsity of statements in which those expressions appear. More recently, Chalmers has extensively theorised the notion of scrutability and set out a number of differing versions of the scrutability thesis (2006; 2012b).

The underlying motivation for the thesis is presumably this: it would be surprising if we were unable to ascertain the referents of our concepts or determine whether statements involving them are true, *no matter how much information we have about the world*. How could *our own concepts* possibly be cognitively closed to us in this manner? So far I have been taking for granted the truth of Chalmers’ scrutability thesis, and its corollary, that materialism requires an a priori entailment of the phenomenal facts by the physical or functional facts. However, this is a topic which will merit further consideration once we turn from the *qualitative* aspect of consciousness to its *subjective* aspect.

[16] One may wonder why phenomenal truths are referred to in connection with the scrutability base for macrophysical truths. The reason is that such truths can relate to heat, colour, and so on, and these arguably have a phenomenal element: for example, knowledge that a given object is a red object includes knowledge about whether it normally causes people to have red experiences (2010, p. 211-212).

4: The subjective/objective distinction

I referred in section 1 to the distinction between the 'qualitative' aspect of an experience, i.e. *what it is like*, and the 'subjective' aspect, i.e. the fact that it is like something *for a particular organism*. Thus far the discussion has centered on the *qualitative* aspect – the 'what it is like'. This follows on from the way in which Chalmers usually frames the problem of accounting for experience – at the start of *The Conscious Mind*, for example, he describes the subject matter of his book as follows:

> We can say that a being is conscious if there is *something it is like* to be that being, to use a phrase made famous by Thomas Nagel. Similarly, a mental state is conscious if there is something it is like to be in that mental state. To put it another way, we can say that a mental state is conscious if it has a *qualitative feel* — an associated quality of experience. These qualitative feels are also known as phenomenal qualities, or *qualia* for short. The problem of explaining these phenomenal qualities is just the problem of explaining consciousness. This is the really hard part of the mind–body problem. (1996, p. 4)

I now want to turn to the *subjective* aspect of consciousness, concentrating on an experience being like something *for a particular organism*. My starting point is a possible challenge to my suggestion that the applicability criteria for phenomenal concepts might be characterizable in functional terms. It could be objected that descriptions such as 'recognizing the colour I term *sky blue*' and 'discriminating the relational aspects of the colour I term *sky blue*' use *phenomenal* senses of 'recognizing' and 'discriminating', whereas what would be required for a materialist explanation of experience is reference to *objective* senses of recognition and discrimination that are describable in purely third-person terms – for example, in terms of observable behaviour. Such an objection acknowledges the possibility of functionally characterized applicability criteria for the concept of a phenomenal quality such as 'sky bluish', but maintains that such criteria cannot sustain a critique of Chalmers' argument because there can be no possibility of any conceptual link between the phenomenal, or first-person, senses of 'recognize' or 'discriminate' and the objective, or third-person, senses of those terms.

This seems, at least on the face of it, an unlikely interpretation of Chalmers' argument. Firstly: if we consider the way in which Chalmers normally sets out the problematic, as in the quote just given from *The Conscious Mind*, he gives a clear impression that the difficulty with conceptually linking phenomenal terms such as 'bluish' and functional terms such as 'recognition' lies with the seemingly ineffable quality of 'bluish', rather than with linking two different notions of 'recognition'. Secondly: it might indeed be true that we have a concept of, say, recognition as a mental activity which we ourselves perform, and *also* a concept of recognition in purely behavioural terms; but it is not immediately clear why there could not be conceptual links between these two different concepts of 'recognition'.

Nevertheless Chalmers does seem to have something like this first-person/third-person distinction in mind. Consider, for example, the following passage from *The Conscious Mind*, describing the underlying intuition on which his anti-materialist argument relies:

> The main intuition at work is that there is something to be explained – some phenomenon associated with *first-person experience* that presents a problem not presented by observation of cognition from the third-person point of view. Given the premise that some explanandum is forced on us by first-person experience that is not forced on us by *third-person observation*, most of the arguments above fall out. It follows immediately, for example, that what needs to be explained cannot be analyzed as the playing of some functional role, for *the latter phenomenon is revealed to us by third-person observation* and is much more straightforward. (1996, p. 110, emphasis added)

Further evidence is provided by the distinction that Chalmers makes (2010, p. 114) between (i) 'a notion of *functional representation,* on which *p* is represented roughly when a system responds to *p* and/or produces behavior appropriate for *p*', and (ii) 'a notion of *phenomenal representation,* on which *p* is represented roughly when a system has a conscious experience as if *p*'. Now this notion of 'functional representation' is explicitly a notion of representational *behaviour*, and by contrasting it with '*phenomenal* representation', it seems that Chalmers is reserving the term 'functional' for objective descriptions in terms of behaviour. On this understanding, functional concepts are purely objective concepts, describable only in third-person terms: hence, our concepts of discrimination, recognition, etc. which we have as a result of our actually performing those mental activities are *not* functional concepts as Chalmers understands the term 'functional'. (It might therefore be helpful to reserve the term 'functional' for descriptions in purely third person terms, and utilize the phrase 'functionally characterized' to cover both first and third-person references to mental processes.)

This implies that the dividing line between the phenomenal and the functional is determined by the distinction between subjective, first-person descriptions and objective, third-person descriptions. Hence, if there is indeed an 'explanatory gap', it is not between phenomenal qualities such as 'bluish' or the smell of a rose on the one hand and mental activities such as recognition or discrimination on the other: it is between the first-person and third-person viewpoints. This poses a difficulty for the argument against materialism, because the idea that there is a conceptual incongruity between these two different senses of a term such as 'recognition' is not as intuitively obvious as the notion that there is a conceptual incongruity between 'recognition' and a qualitative term such as 'bluish'. If the anti-materialist is to convert this difference between first-person and third-person descriptions into a convincing argument against materialism then they will have to provide a strong reason as to *why* we cannot conceptually link these two different sorts of description of functionally characterized mental processes.

In fact, there does seem to be such a reason, but it is a two-edged sword so far as the anti-materialist is concerned. The objective, third-person viewpoint derives from the possibility of *inter-subject*

agreement on facts, where that agreement can be obtained, at least in principle, from *any* person placed in the relevant situation. Hence, the third-person viewpoint is not to be thought of as a perspective on some objective realm, separate from another realm which is accessible only from a privileged first-person viewpoint (a conception which would simply reinstate the dualist assumption which Chalmers' argument strives to avoid); it is instead a *generalization* of the first-person viewpoint. Now as Chalmers notes (e.g. 2010, p. 15), scientific investigation of phenomena in the physical world uses descriptions that can be put in terms of structure and dynamics. We can surmise that the reason why this is the case is that descriptions in those terms are capable of generating inter-subject agreement that is strong enough for an objective, scientific fact, being based upon *agreed measurement criteria* for observations concerning space (giving us objective *structure*) and time (giving us objective *dynamics*). However, this type of inter-subject agreement does not seem possible in respect of phenomenal facts: each experience is only had by a single person, and phenomenal facts are therefore not cognitively available to anyone other than the individual who is having the relevant experience. All we can have in these instances is inter-subject agreement on the *behaviour* of people who are subject to certain types of stimuli, including their expressed judgments about how their experiences feel to them. So now we *do* have a reason why it is not possible for there to be a priori entailments from physical or functional facts to phenomenal facts: namely, that there cannot be objective applicability criteria, available in principle to any third party, for the use of a concept of a phenomenal aspect of an individual's experience.

However, what guarantees this lack of a priori entailment is the 'interiority' of experience – what Tom Clark refers to (2005; 2010) as an experience's 'privacy' – and we have at least plausible accounts of the development of interiority as a mental characteristic of higher organisms, e.g. Damasio, 2010; Deacon, 2012; Humphrey, 1992; Thompson, 2007; Varela, 1997. It is true that anti-materialists have criticised the likes of Humphrey for not providing any explanation of the subjectivity or interiority of *experiences*. But my discussion here concerns the applicability criteria for the use of concepts: hence, all that needs to be explained is how the evolutionary development of organisms, and of their nervous systems and sensory apparatus, leads to the constitution of both an interiority and an exteriority with respect to that interiority, such that some of the phenomena of which higher organisms are aware are capable of being the objects of inter-subject agreement, whilst other phenomena of which they are aware are not capable of producing such inter-subject agreement.

At some point in the future therefore, we might conceivably have: (i) an explanation as to how a first-person viewpoint, *modulo* experience, can come to exist in higher organisms; (ii) explanations of the qualitative aspects of experiences in terms of our own first-person mental activities such as, say, discrimination and recognition; and (iii) explanations in neurophysiological terms as to how functions such as discrimination and recognition are instantiated in the brains of higher organisms. Chalmers' argument would maintain that we must nevertheless rule out materialism, in favour of dualism or

perhaps panpsychism, because the applicability criteria for our concepts of discrimination and recognition that feature in (ii) cannot be expressed using our concepts of discrimination and recognition that feature in (iii), for reasons that are inherent in the notion of a first person point-of-view, the existence of which has been explained in (i). This is the sort of argument which could only convince those who are *already* convinced, and is a long way indeed from Chalmers' argument as it is usually understood. For example, it is difficult to see how Chalmers could maintain that an identity claim between the functions that feature in (ii) and the functions that feature in (iii) must be 'brute' or 'explanatorily primitive' in the way that he does for claims about the identity of experiences and physical processes (e.g. 1997, pp. 12-16).

Still, one would have to accept that there could be no a priori entailment of phenomenal facts by facts wholly expressed in objective, third-person terms. But if the reason for this is simply that an actual experience is only ever available to a single person, thereby precluding inter-subject applicability criteria for our phenomenal concepts, then we have good reason to question Chalmers' thesis concerning scrutability. As we have seen, Chalmers' view is that scrutability applies to almost all our judgments about the way things are in the world. However, interiority and the difference between first-person and third-person viewpoints give us a good reason for thinking that phenomenal concepts are indeed unusual: their applicability criteria are only cognitively available to the person having the relevant experience – only *you* can determine whether a concept of 'bluish' applies to your current experience. Hence those criteria cannot be expressed in terms of agreed measurements of space (i.e. objective structure) or of time (i.e. objective dynamics): that is, they cannot be expressed in purely physical terms. The referents of our phenomenal concepts might actually *be* higher-level, psychologically constituted properties of brains, but we could never be in a position to know this a priori. Once again, Chalmers' argument against materialism would fail for the reason put forward by type-B materialists: we would have a clear epistemological explanation for the lack of a priori entailment between physical or functional facts and phenomenal facts, and hence no reason to believe that the epistemological gap necessarily implies an ontological gap.

5: The arguments in *The Conscious Mind*

I have now examined the 'master version' of Chalmers' argument in terms of both the qualitative and the subjective aspects of experience. This is therefore an appropriate point to consider what the discussion thus far has to tell us about Chalmers' case against materialism as he sets it out in *The Conscious Mind*. In that book, he presents five specific arguments purporting to show that experience cannot be logically supervenient on the physical or functional, and hence that a reductive explanation of experience must be impossible:

1. The conceivability of zombies.

2. The conceivability of an inverted spectrum.

3. ‘Mary’ learns new knowledge on seeing a red object.

4. Our knowledge about phenomenal experience is obtained mainly from our ourselves, rather than from the external world.

5. The impossibility of analyzing phenomenal experience in terms of anything else.

As I noted earlier, Chalmers has said (2010, p. xv) that the argument put forward in ‘Facing Up’ is more fundamental than the thought experiments in *The Conscious Mind*, since the latter turn on points concerning structure and function which are similar to those made in his earlier article. In fact the general form of the argument in ‘Facing Up’, together with the more specific arguments in *The Conscious Mind*, provide what Chalmers’ refers to (1999) as a ‘circle of mutual support’. The claims that we can conceive of inverted qualia and zombies, that Mary learns new knowledge when she leaves her black-and-white room, and that we cannot analyse phenomenal experience or learn about it from the outside world – these all provide support for the thesis that there is no a priori entailment between physical or functional facts and phenomenal facts. If there *were* such entailment, then our attempts to conceive of zombies or inverted qualia should reveal some logical contradiction, Mary should be able to deduce all relevant facts before she experiences seeing a red object, and we should be able to arrive at some analysis of the qualitative aspect of an experience in physical or functional terms and thereby learn about it from external observation. That we cannot do so supports the argument in ‘Facing Up’ that physical structure and function cannot explain conscious experience; and that argument is supposed to help us to understand exactly *why* zombies and inverted qualia are conceivable, *why* it is that Mary learns something new, and *why* experience is unanalysable and can only be learnt about from our own personal case.

So how does the discussion in sections 3 and 4 impact on these specific arguments in *The Conscious Mind*? If, as I suggested in section 4, there cannot be objective, third-person applicability criteria for the use of concepts of phenomenal properties, and therefore there can be no possibility of any a priori entailment of phenomenal facts by physical or functional facts, then we would of course expect the specific arguments to reflect that lack of entailment. Nevertheless, it is instructive to take a closer look at those five arguments – and the three thought experiments in particular – in the context of my discussion of the ‘master version’ of Chalmers’ argument.

(a) Zombies

Let us start with the case of zombies. John Searle helpfully suggests that we consider this type of thought experiment from the first-person perspective: *what would it be like for me?* I can envisage that my external behaviour and everything observable about my brain remains as it is now, but that I

slowly lose the light of conscious experience – I gradually become a zombie (Searle, 1992, pp. 66-67). Why does this seem to contain no logical inconsistency?

The earlier discussion suggests two reasons. Firstly, when I think about my experiences I usually think of the qualitative aspects of those experiences, such as colours, smells, and so on. So even if it is the case that those qualitative aspects are conceptually reducible to my own mental activities, such as recognition or discrimination, I nevertheless tend to characterize those qualitative aspects of experiences without using functional terms, e.g. as 'bluish'. The effect is one of distancing my intuitive appreciation of my experiences from their underlying ground in my own mental processes – a ground which could be revealed by a critical investigation of the epistemic intensions of the relevant phenomenal concepts. The second reason is that there is indeed an unbridgeable epistemic gap between those mental processes as I am aware of myself performing them and their objective manifestations in terms of externally observable behaviour. Such a gap exists for two reasons – firstly, the interiority of consciousness means that there can be no objective, inter-subject applicability criteria for our concepts of our own psychological processes as we ourselves perform them; and secondly, it is difficult to see how any set of objective facts, observable by any suitably-placed third-party, could preempt the first-person criteria which I have for my concepts of my *own* mental activities as I experience them.

The result is that when I try and conceive of myself as a zombie, what I bring to mind is objectively describable behaviour and brain activity on the one hand, and my own experiences (or a lack thereof) characterized in purely qualitative terms such as 'bluish' on the other. But the latter are 'doubly distanced' from the former: not only is there an epistemic gap between my two notions of psychological terms such as recognition or discrimination, but I also intuitively think of the qualitative aspects of my experiences in terms of their *characterizations* rather than in terms of the *applicability criteria* for my concepts of those qualitative aspects. Hence I can perceive no contradiction between my being physically and behaviourally as I am now, yet without any experiences.

(b) Inverted spectrum

The case of inverted qualia is broadly similar. People tend to perceive no logical inconsistency in imagining that they might have had what they would describe as a 'bluish' experience rather than a 'reddish' experience, even if all physical and functional facts remain identical. The reasons for this are the same as in the case of imagining a zombie: (i) the distancing of our intuitive characterization of experiences (e.g. as 'bluish' or 'reddish') from their underlying ground in our mental processes, and (ii) the epistemic gap between those mental processes as we are aware of ourselves performing them and their objective manifestations in terms of externally observable behaviour. This 'double-distancing' leads us to think of the phenomenal aspects of experiences as *intrinsic* properties – as

being simply ‘bluish’ or ‘reddish’ – and it is these supposed intrinsic properties which we conceive of as being inverted with respect to our mental activities.

Those mental activities include the discriminations and comparisons we make which constitute the relational structure that our experiences bear one-to-another. Take the example of the relational structure represented by a *colour space* (Clark, 1993; Hardin, 1988) – it seems to many people to be conceivable that the colours which we experience could be inverted with respect to each other (along some axis of the colour space) whilst maintaining the same relational structure. There is some doubt as to whether this is true in respect of human colour space, due to its asymmetries (see Byrne, 2010), but in any event it is conceivable that there could be symmetric quality spaces which would raise the same difficulty.[17] Chalmers, for example (1996, p. 101), hypothesizes a creature which can experience only two colours, such that the distinction between them represents the whole of its colour space; it is logically conceivable, he claims, that two such creatures might share exactly the same physical and functional properties, yet one experiences ‘colour A’ whilst the other experiences ‘colour B’. However, once we exclude intrinsic phenomenal properties, it becomes difficult to conceive how this could be the case. With such a simple colour space, a creature could only conceptualize its experiences of one of the colours as, say, ‘the *one* colour’ and its experiences of the other colour as ‘the *other* colour’. In the absence of intrinsic phenomenal properties, there is nothing to be inverted: there is no way in which the ‘the *one* colour’ could be ‘phenomenal quality B’ rather than ‘phenomenal quality A’ – it is just ‘the colour I recognize as the *one* colour’, rather than ‘the colour I recognize as the *other* colour’. The same holds true in the vastly richer human colour space: in the absence of intrinsic phenomenal properties there is no conceptual room for me to experience a colour that I recognize as that which I term ‘sky blue’, and which has those relations to my other experiences of colour that sky blue has, and yet for me to be actually experiencing a very dark red. The relations of colour space, together with re-identification of the sky bluish aspect of my experience, are sufficient to individuate and categorize colour experiences. Hence, it is plausible that they also characterize the applicability criteria for my concept of the sky bluish aspect of my experiences, and if those criteria are met then there is no conceptual possibility that I could be experiencing a colour other than sky blue. The impression that I could do so is due to the ‘double-distancing’ of our intuitive characterizations of phenomenal qualities from our concepts of objectively described mental processes.

(c) The knowledge argument

The case of the knowledge argument is a little more complicated. In Frank Jackson’s famous thought experiment (1982; 1986), Mary is a super-scientist who knows all of the physical and functional facts concerning vision, colours, the brain, and so on. Nevertheless, says Jackson, it seems that she will

[17] It has been argued, for example by Dennett (1993) and by Hilbert & Kalderon (2000), that quality spaces *must* be asymmetric, but I will ignore that possibility here.

learn various facts about phenomenal experience after she leaves the black-and-white room in which she has lived all her life – facts that she could not have previously known, or deduced, no matter how complete her scientific knowledge. Chalmers argues (1996, pp. 103-104) that Jackson's thought experiment shows that phenomenal facts cannot logically supervene on the physical or functional facts, and hence that reductive explanation of experience must fail.

It is tempting to suppose that what Mary learns when she leaves her black-and-white room and first sees a red object is along the lines of 'what it is like to experience red', but this formulation is problematic. Firstly, the aim of the knowledge argument is to prove that phenomenal facts do not follow of necessity from the physical facts and it is difficult to see how this can be done unless the relevant knowledge is propositional in form, since necessity is a relationship between propositions (Rabin, 2011). There are, however, reasons to doubt whether 'knowing what it is like' can be expressed in propositional terms (Nida-Rümelin, 1995). Secondly, Jackson himself argued (1986) that the facts of which Mary lacks knowledge are not those relating to her own experiences of colour – since there are no such experiences, and hence no such facts, while she is still in the black-and-white room – but facts concerning the world outside, such as the nature of *other* people's experiences when they see a coloured object. For my purposes here, I shall simply take it that for Mary to have new knowledge there must be *a reduction in the epistemic possibilities* (Lewis, 1988; Nida-Rümelin, 1995). This may be as straightforward as learning that experiencing seeing a red object is like *this*, whereas it might have been like *that* (Chalmers, 1996, pp. 144-145).

We can illustrate this requirement by supposing that while Mary is still restricted to her black-and-white room, she is presented with a series of coloured squares of paper but given no information about what colours they actually are (nor is she allowed to establish the identity of each square by scientific experiment).[18] Mary could be looking at a blue piece of paper but still not know that when people look at the sky they experience this same colour. She does not learn *that* until she sees the sky herself or learns that the square of paper is 'blue'. Up to that point, she cannot rule out the possibility that other people experience the colour that is on another piece of paper that she has seen. Once Mary has seen the sky she can exclude those other possibilities, and thereby learn a new fact about the world (Nida-Rümelin, 1995) – or so the proponent of the knowledge argument maintains.

This requirement that epistemic possibilities be excluded means that Mary's new knowledge must relate to some property of her experience which enables her to distinguish it from other experiences. For the purposes of the knowledge argument, this cannot be a property which Mary could learn about from her extensive knowledge of physical and scientific fact. It must therefore relate to the

[18] My example of Mary being presented with coloured squares of paper, but not knowing which colour they are, is similar to Nida-Rümelin's (1995) variation of Jackson's thought experiment. However, Nida-Rümelin's purposes differ substantially from mine, one of hers being to defend the knowledge argument from the objection by Lewis (1988) that all Mary learns is a set of abilities. Papineau (2002) uses an example of Mary being shown a piece of paper of an unknown colour in order to show that she cannot relate this experience to her existing physical concepts.

phenomenal aspect of the experience *as such*, and it is difficult to see what else that aspect could be other than some intrinsic phenomenal property that belongs to the experience. In which case, we may suppose that the *reason* why Mary is able to learn that when other people look at the sky they experience the same colour that she did when she looked at her unknown square of paper, is that both experiences have the same intrinsic property – phenomenal blueness, say, rather than phenomenal redness.

However, as I discussed in section 3, Chalmers' case against materialism needs to avoid any question-begging assumption concerning the existence of intrinsic phenomenal properties. So what happens if we attempt to reframe our description of Mary's situation so that we do not smuggle into the thought experiment some assumption concerning the existence of such properties? Suppose that Mary is shown two coloured squares of paper and is informed about their respective colours: 'this is red', 'this is green'. Then she is shown a square with another colour but is not told what it is: she can see that it isn't red and it isn't green, but other than that her experience is just of a colour she hasn't seen before.[19] The proponent of the knowledge argument wants to maintain that there is nevertheless a fact about *which* colour Mary is experiencing – for example, that it is the same phenomenal quality that *your* experience has when you look at the sky on a clear day; if there is no such fact, then there is no possibility of her experience having an intrinsic property such as phenomenal blueness or phenomenal redness. I suggest that we can deny that there is any such fact, and therefore that there is any intrinsic phenomenal property belonging to Mary's experience, on the grounds that *the phenomenal facts should be determined by the phenomenology of the experience*: Mary is having an experience of a colour which isn't red or green and which she doesn't recognize, and this exhausts the experience's phenomenal characteristics. But if there is no intrinsic property belonging to *this* experience, then we have a good reason for thinking that there are no intrinsic properties belonging to *other* experiences, such as those of normal observers when they observe a blue object. In which case, there can be no new phenomenal fact relating to such properties for Mary to learn.

The point of this reformulation of Mary's situation is to show that the knowledge argument gains its credibility not just from the aforementioned double-dissociation of the qualitative aspects of experiences from objectively describable mental activities, but also from assuming at the outset that there is some intrinsic phenomenal property which is common to all experiences that people have on seeing (say) a blue object under normal conditions – a property that is common to my experience, to your experience, and to Mary's experience when she first sees such an object. But the situation is different if we understand the qualitative aspects of experiences along the lines suggested in section 3: for example, that I apply my concept of *sky blue* to my experiences if I can *recognize* the colour that I am experiencing as that which I term 'sky blue' and if I can *discriminate its relational properties* with

[19] I am here ignoring the possibility that Mary can deduce something about the colours that she experiences from their relations with each other, as formalized in models of colour space (Hardin, 1988).

respect to other colours as being those which I understand 'sky blue' to have. There is now no sense in which there can be a common property shared by my experience, your experience, and Mary's experience: the phenomenal property of *my* experience when I see a sky blue object ('a colour which I, Mike, recognize as that which I term *sky blue*') differs from the property of *your* experience when you see the same object ('a colour which I, the Reader, recognize as that which I term *sky blue*'), and from the property of Mary's experience when *she* observes a sky blue object for the first time in her black-and-white room ('a colour which I, Mary, do not recognize'): the phenomenal properties reflect the subjective nature of the experiences of separate individuals. The alternative is to maintain that phenomenal properties are in some sense objective: that if you, and I, and Mary in her black-and-white room, all look at some sky blue object, then our experiences will all have the *same* phenomenal property. This would be to treat having an experience in the same manner as perceiving an object: for example, if you and I and Mary were all looking at a mountain, it would be the *same* mountain we were looking at even if we all perceived it from completely different angles and distances. But the mountain is an object, and experiences are not.

Descriptions of phenomenal properties in functionally characterized terms will therefore vary with the individual's situation, whereas phenomenal properties expressed in intrinsic terms represent a strong ontological intuition in respect of the nature of phenomenal qualities. Just how strong that intuition is can be seen from Chalmers' description of the epistemic implications: 'phenomenal knowledge … is a sort of *objective* knowledge of the world, *not essentially tied to any viewpoint*' (2010, p. 163, emphasis added). It might be thought extraordinary that this is where Chalmers has ended up, from a starting position in 'Facing Up' where he insisted on experience as the ineliminable '*subjective* aspect' of consciousness (2010, p. 5): something seems to have gone very wrong somewhere.

(d) Epistemic asymmetry

Chalmers maintains that the only knowledge that we have of conscious experiences derives from our own personal access to them, rather than from knowledge of the physical world. This implies that facts about such experiences cannot be logically supervenient on the physical or functional facts (since if they *were* supervenient, we would have another route by which we could gain knowledge about our experiences):

> [C]onsciousness is a surprising feature of the universe. Our grounds for belief in consciousness derive solely from our own experience of it. Even if we knew every last detail about the physics of the universe—the configuration, causation, and evolution among all the fields and particles in the spatiotemporal manifold—*that* information would not lead us to postulate the existence of conscious experience. My knowledge of consciousness, in the first instance, comes from my own case, not from any external observation. It is my first-person experience of consciousness that forces the problem on me. (1996, pp. 101-102)

We can view this ‘epistemic asymmetry’ as a straightforward result of the interiority, or privacy, of consciousness. Experiences are interior phenomena that are only cognitively available to a single organism, rather than objects or processes concerning which there can be inter-subject agreement and therefore objective knowledge.

(e) *Absence of analysis*

Chalmers’ discussion of our supposed inability to analyse conscious experiences in physical or functional terms (1996, pp. 104-106) – which includes his comment that ‘[t]o analyze consciousness in terms of some functional notion is either to change the subject or to define away the problem’ – has already been outlined in section 2, and much of the discussion in sections 3 and 4 is directly relevant to it. No more needs to be added here.

The specific arguments in *The Conscious Mind* display three important aspects of Chalmers’ case against materialism, all of which have already been noted in sections 3 and 4. Firstly, they reflect the impossibility of there being third-person, objective applicability criteria for a concept relating to an aspect of an experience that is only cognitively available to the person who has that experience. However, they also display two of the weaknesses of Chalmers’ case against materialism:

- His failure to distinguish our intuitive characterization of the qualitative aspect of an experience (e.g. ‘bluish’) from the applicability criteria for our concept of that qualitative aspect.
- His tendency to implicitly assume the existence of intrinsic phenomenal properties – something which is evident in both the ‘inverted qualia’ and ‘Mary’ thought experiments.

6: Chalmers and the phenomenal concept strategy

As well as claiming that the epistemic gaps which relate to consciousness imply the falsity of materialism, Chalmers has put forward a separate argument (2010, Chapter 10) which is intended to show that any attempt to provide a *physical explanation for those epistemic gaps* must fail. That argument is aimed at what is referred to as the ‘phenomenal concept strategy’ (henceforth ‘PCS’) which maintains that such gaps are due to the special nature of our phenomenal concepts – for example, that they are *recognitional* or *quotational* concepts. However, Chalmers frames his argument quite generally, in terms of any ‘psychological feature’ which might explain the epistemic gaps. So I therefore need to consider the possible impact of his argument against the PCS on the proposal which I made in section 4: that the epistemic gaps might be due to the impossibility of third-person applicability criteria for phenomenal concepts – something which is physically explicable in terms of the interiority or privacy of consciousness.

(a) Chalmers' argument against the PCS

If the PCS succeeds, says Chalmers, then 'we may not have a straightforward physical explanation of consciousness, but we have the next best thing: a physical explanation of why we find an explanatory gap' (2010, p. 305). His argument that even this 'next-best' explanation cannot succeed starts by posing a dilemma: if P is the totality of microphysical facts, and C is the relevant 'psychological feature' that is put forward as an explanation of the epistemic gaps, then either (i) P&-C is conceivable, or (ii) P&-C is not conceivable. Now as we saw earlier, Chalmers understands 'explanation' in terms of a priori entailment, such that if X&-Y is conceivable, then there can be no a priori entailment – and hence no explanation – from X to Y. Therefore, taking each horn of the dilemma in turn:

> (i) If P&-C is conceivable, then the physical facts cannot explain the relevant psychological feature.
>
> (ii) If P&-C is *not* conceivable, then a zombie (which is identical to a human in all physical respects) must have C. But *our* epistemic situation (call this 'E') clearly differs from that of a zombie: for example, when a zombie version of Mary – the super-scientist who has not yet experienced colours – views a red object for the first time, she will not gain any new knowledge, whereas the human Mary *will* gain new knowledge.[20] Hence, having C is not sufficient to explain E, and therefore C cannot explain the epistemic gaps which we encounter.

The result, claims Chalmers, is that the PCS only leads to a further epistemic gap: either between P and C (i.e. the special psychological feature is not in fact explainable in physical terms), or between C and E (i.e. the psychological feature cannot explain the epistemic gaps that we find between physical and phenomenal facts). Either way, the PCS must fail.

(b) What does Chalmers mean by 'epistemic situation'?

One aspect of this argument which gives pause for thought is Chalmers' use of the term 'epistemic situation'. Initially, he uses this phrase in the context of explaining the epistemic gaps between the physical and the phenomenal; for example, he refers to the claim that 'C explains our epistemic situation with regard to consciousness (C explains why we are confronted with the relevant distinctive epistemic gaps)' (2010, p. 311). But he then goes on to *define* 'epistemic situation' in terms of the truth values of a subject's beliefs and whether those beliefs are justified or cognitively significant. For example, both Mary and her zombie twin share the same belief – that they have gained new knowledge on seeing a red object – but only Mary's belief is a *true* belief; hence, Mary shares in our

[20] Or so Chalmers claims. In the previous section I argued that Mary actually learns no new propositional fact, but for the purposes of describing and evaluating Chalmers' argument against the PCS I shall follow him in taking it that Mary can learn some new fact that her zombie-twin cannot learn.

epistemic situation, but her zombie twin does not. Chalmers' justification for defining 'epistemic situation' in terms of the truth and justifiability of beliefs is that the distinctive epistemic gaps which the PCS attempts to explain are themselves framed in terms of true beliefs and knowledge:

> In the anti-physicalist's arguments, the relevant epistemic gap (from which an ontological gap is inferred) is characterized in such a way that truth and knowledge are essential. For example, it is crucial to the knowledge argument that Mary gains new factual *knowledge* or at least new true beliefs. It is crucial to the conceivability argument that one can conceive beings that lack phenomenal states that one actually has. ... If one characterized these gaps in a way that were neutral on the truth of phenomenal beliefs, the arguments would not get off the ground. So truth-value is essential to the relevant epistemic gaps. If so, then to undercut the inference from these gaps to an ontological gap, the phenomenal concept strategy needs to show how the relevant truth-involving epistemic gaps are consistent with physicalism. (2010, p. 324)

However, while a materialist might accept that what needs to be explained is the existence of the *actual* epistemic gaps that we encounter, they might object that the truth or justification of the beliefs associated with those gaps depends directly upon the existence of the relevant phenomenal properties. But in the case of a zombie those properties are absent *by definition*; and therefore Chalmers' argument against the PCS would appear to beg the question.

Chalmers' response is that he can circumvent any accusation of begging the question by characterizing both the special psychological feature and our epistemic situation in *topic-neutral* terms, i.e. in terms that do not explicitly attribute phenomenal states or concepts. So whereas Mary might be in a *phenomenal* state, her zombie twin can be described as being in a *schmenomenal* state (2010, p. 326); and we can refer to both types of states using *quasi-phenomenal* concepts – which, Chalmers suggests, might be understood as concepts associated with various types of psychological processes:

> The restriction to topic-neutral terms allows that thesis *C* may include psychological or epistemological vocabulary, but phenomenal vocabulary is barred. For example, instead of casting thesis C as a thesis explicitly about phenomenal concepts, one can cast it as a thesis about *quasi-phenomenal* concepts, where these can be understood as concepts deployed in certain circumstances that are associated with certain sorts of perceptual and introspective processes and so on. Phenomenal concepts will be quasi-phenomenal concepts, but now it is not out of the question that zombies might have quasi-phenomenal concepts too. (2010, p. 314)

This, he claims, can deal with the circularity argument:

> [It has been claimed] that if we characterize the epistemic gap as requiring the truth of phenomenal beliefs, then we will illegitimately build phenomenology into the epistemic gap by definition. However, ... the epistemic gap can be characterized topic neutrally, perhaps along the following lines: we possess a quasi-phenomenal concept *q,* such that our quasi-phenomenal belief *someone has q* is true ... and *if P then someone has q* is not a priori [where P is the totality of microphysical facts]. This

> characterization of the epistemic gap requires truth of the relevant belief and does not build in any claims about phenomenology. If the existence of an epistemic gap so construed could be physically explained, this could reasonably be taken to undermine arguments from an epistemic gap to an ontological gap. (2010, p. 325)

However, two problems arise here. Firstly: if we are restricted to non-phenomenal terminology, then it would seem that my epistemic situation, *so characterized*, cannot possibly diverge from that of my zombie equivalent – after all, I am identical with my zombie twin in all non-phenomenal respects and therefore any non-phenomenal description of the zombie must apply equally to me.[21] In which case, there is no difference in our epistemic situations, and Chalmers is not entitled to his conclusion that C cannot explain E. Secondly: it is difficult to see how Chalmers' criterion – '*if P then someone has q* is not a priori' – can possibly be met in the case of a concept of a schmenomenal state. The physical or functional nature of the applicability criteria for this type of non-phenomenal concept suggests that the fact that someone is in a particular schmenomenal state should always be derivable a priori from the totality of microphysical facts.

(c) Critiquing Chalmers' argument against the PCS

There are, therefore, a number of ways in which a critic might approach Chalmers' argument against the PCS. They can dispute his conclusion in respect of either, or both, horns of his dilemma; they can question his use of the notion of 'epistemic situation'; or they might query the appropriateness, or results, of using topic-neutral terminology. The main options taken by the critics are:

(1) Claim that the PCS *explains any new epistemic gap just as much as the original epistemic gap* – or, alternatively, that the new gap is just the old gap re-described (Balog, 2012; Elpidorou, 2013; Papineau, 2007)

(2) Argue that *zombies do in fact share our epistemic situation* as Chalmers defines it (Carruthers & Veillet, 2007; Elpidorou, 2013; Papineau, 2007)

(3) Maintain that the PCS *only has to explain the existence of epistemic gaps*, not our entire epistemic situation (Diaz-Leon, 2010)

In fact, there is something rather strange about Chalmers' entire case against the PCS. Recall that conceivability arguments are simply intended to disclose any lack of a priori entailment between

---

[21] One might wonder just how to characterise a zombie's epistemic situation. However, Elpidorou points out (2013) that one can evaluate Chalmers' critique of the PCS without having to say anything substantive about what might constitute a zombie's epistemic situation. If it is possible to give an appropriate description of *my* epistemic situation in topic-neutral terms (i.e. eschewing all phenomenal terminology), then that description must also apply to my zombie-twin – since we are completely identical in all respects bar the phenomenal – and this is what matters so far as Chalmers' argument against the PCS is concerned. If, on the other hand, such a topic-neutral description is *not* possible, then characterising my epistemic situation would require the use of phenomenal concepts – which the zombie lacks by definition – and Chalmers' argument against the PCS would then beg the question. Either way, the exact nature of the zombie's epistemic situation is of no significance.

different descriptions. The same is true of epistemological arguments based on our knowledge of phenomenal facts. In *The Conscious Mind*, for example, the conceivability of zombies, the conceivability of inverted qualia, and Mary's new knowledge on first seeing a red object, are all invoked as evidence that the phenomenal facts do not logically supervene on the physical or functional facts, and therefore cannot be a priori entailed by them. Chalmers actually prefaces his discussion of these thought experiments by saying:

> How can we argue that consciousness is not logically supervenient on the physical? There are various ways. We can think about what is conceivable, in order to argue directly for the logical possibility of a situation in which the physical facts are the same but the facts about experience are different. We can appeal to epistemology, arguing that the right sort of link between knowledge of physical facts and knowledge of consciousness is absent. (1996, pp. 93-94)

So if we can provide a physical explanation of the conceptual independence of phenomenal concepts – that is, we can provide a physical reason why there cannot be a priori entailments between those concepts and non-phenomenal concepts – then that should be that: what more could be asked of the materialist? Yet Chalmers does seems to want more, because he specifically rejects the suggestion that it is sufficient that:

> a physically explicable account of phenomenal concepts can explain ... the *inferential disconnection* between our physical and phenomenal beliefs, including the fact that the latter are not deducible from the former ... (2010, pp. 323-324)

(d) What is the aim of Chalmers' argument against the PCS?

In order to make sense of Chalmers' stance here, we have to understand how his argument against the PCS fits into the overall dialectic of his dispute with type-B materialists. Chalmers' general case against type-B materialism (e.g. 2010, pp. 115-118) is that it relies upon either the assertion of a brute and unsupported identity, or some unjustifiable and unnecessary form of modality. So when he sets out to consider the PCS, Chalmers envisages that this particular variant on a type-B position will have to amount to *more than* a claim that the nature of our phenomenal concepts results in an inferential disconnect from the physical facts. After all, both Chalmers and the type-B materialists accept that our phenomenal concepts are such that the physical facts do not entail the phenomenal facts; the dispute between them is whether the resulting epistemic gaps must imply an ontological gap. Chalmers is therefore careful to explain that the aim of his critique of the PCS is *not* to establish the impossibility of type-B materialism:

> ... I am not arguing in this chapter that type-B materialism is false. I do that elsewhere ... Here I am simply arguing that the phenomenal concept strategy provides no support for type-B materialism and *provides no grounds for rejecting arguments from the epistemic gap to an ontological gap*. (2010, p. 322, emphasis added)

If they are to make any headway, proponents of the PCS must therefore attempt to *justify* the claim that the epistemic gaps which we come across in respect of consciousness are somehow consistent with materialism, notwithstanding the general case which Chalmers has already put forward against such a position. Chalmers suggests that they might attempt to do this by proving that such gaps can arise in a purely physical world – because if that were the case, it would directly imply the failure of the anti-materialist's inference from an epistemic gap to an ontological gap.

Of course, Chalmers anticipates the complaint that he is holding the PCS to too strong an obligation. However, he insists that a weaker requirement would mean that the PCS would add nothing to what type-B materialists already claim to be the case:

> [The PCS] proponent may respond by saying that ontological physicalism is compatible with the existence of explanatory gaps. *But now we are back where we started, before the phenomenal concept strategy came in.* Antiphysicalists argue from an epistemic gap to an ontological gap. The phenomenal concept strategy ... was supposed to *ground the rejection of this inference by showing how such epistemic gaps can arise in a purely physical system*. If successful, the strategy would help to justify the claim that the epistemic gap is compatible with ontological physicalism and so would lend significant support to type-B materialism. (2010, pp. 321-322, emphasis added)

Chalmers' argument against the PCS is therefore intended to prove that no version of that strategy can possibly show that the type of epistemic gap which we find in respect of consciousness can arise in a wholly physical system. And if it turns out that the PCS actually amounts to no more than a description of some psychological feature which might explain the inferential disconnect, then it does nothing to advance the type-B materialist's position.

What Chalmers is looking for from his critics might therefore be summarised as follows: 'Give me *either* a good reason why my earlier arguments – to the effect that the epistemic gaps which we find in respect of consciousness must imply an ontological gap – are wrong, *or* provide something that shows that the same sort of epistemic gaps could arise in a system that is wholly physical'. This is, one suspects, the reason why he appears unconcerned by the claims of Balog (2012), Elpidorou (2013), and Papineau (2007) that the PCS can explain the epistemic gap between P and C, or between C and E, in the same way that it can 'explain' the gap between the physical facts and the phenomenal facts – to claim this is, he believes, to add nothing of substance to what type-B materialists already maintain is the case.

(e) Chalmers and type-B materialism

What this means is that I cannot consider Chalmers' argument against the PCS in isolation from his more general reasons for rejecting type-B materialism – and it is to those reasons that I now turn.

Type-B materialism combines an acceptance of the epistemic gaps between the physical and the phenomenal with a rejection of any inference to an ontological gap (2010, p. 115). Sometimes the

position is characterized as the view that zombies or inverted qualia are not *metaphysically* possible, even if they are *logically* possible (1996, p. 131). Chalmers has set out his objections to type-B materialism on a number of occasions (1996, pp. 131-140;[22] 1997, pp. 11-16; 2010, pp. 115-118, 165-192), and what underlies his rejection of this position is the following consideration: once we accept that there are two separate phenomena which cannot be linked in an a priori manner, it is difficult to see how a claim that those phenomena are identical (e.g. that the bluish aspect of my experience actually *is* the firing of certain neurons in my brain) can be justified. The two phenomena represent *different* properties or states, irreducible to each other – so what possible justification could there be for asserting that they are in fact *identical*?

Even those a posteriori identities which we discover through empirical discoveries – such as 'water = H_2O' or 'Samuel Clemens is Mark Twain' – could, in principle, be arrived at *a priori* by someone who knew all the physical facts and understood the relevant concepts (1997, p. 13). Conscious experience is, however, completely different: in this case, there is a gap in a priori terms between experience and the *totality* of physical facts. Once one accepts – as the type-B materialist does – the existence of this sort of epistemic gap between the two types of phenomena, the assertion that those phenomena are somehow identical would seem unwarranted and unmotivated – except by a desire to save materialism:

> The problems [of type-B materialism] are all rooted in the same place: it makes the identity an *explanatorily primitive* fact about the world. That is, the fact that certain physical/functional states are conscious states is taken as a brute fact about nature, not itself to be further explained. (1997, p. 12)

Chalmers backs up this objection by questioning the very notion of metaphysical necessity that is often used to support the assertion that materialism might still be true despite the epistemic gaps. Firstly, he argues in favour of what he terms modal rationalism and the existence of strong links between the epistemic, modal and ontological realms (2002b; 2010, pp. 184-192). Secondly, he argues that modal rationalism implies that there are no 'strong metaphysical necessities' (1996, pp. 136-138), and provides objections against a number of proposed examples of such necessities (2010, pp. 170-180). Thirdly, he claims (1996, pp. 131 -134; 1997, p. 14) that the whole notion of *metaphysical* necessity – a modal concept which is distinct from both logical necessity and natural necessity – is largely the result of a misreading of Kripke's treatment of a posteriori necessity, and that we therefore have no reason to believe that such a mode of necessity exists.

Chalmers is blunt in his choice of words when he argues that the materialist's reliance on some form of identity or metaphysical necessity is without motivation or justification. This is, he says, 'more

[22] At this point in *The Conscious Mind*, Chalmers has yet to introduce the term 'type-B materialism', and instead refers to materialists who rely on some form of a posteriori necessity.

akin to theft than to honest toil' (2010, p. 116), 'at best ad hoc and mysterious and at worst incoherent' (2010, p. 117), and 'brute and inexplicable' (1996, p. 137). In summary:

> [T]he problems of type-B materialism can be expressed both on intuitive and technical grounds. On the most intuitive grounds: it is a solution by stipulation, which 'solves' the problem only by asserting that brain states are conscious states, without explaining how this can be. On slightly more technical grounds: it requires an appeal to a primitive axiom identifying consciousness with a physical process, where this identity is not derivable from the physical facts and is thus unlike any identity statement found elsewhere. On the most technical grounds: it either rests on an invalid appeal to Kripke's a posteriori necessity or requires a new and stronger notion of metaphysical necessity in which there is no reason to believe. (1997, p. 15)

Given that Chalmers has already set out this general case against type-B materialism, the problem faced by the PCS is that it needs to show *how* something can be conceivable, *while still being impossible* (2010, p. 182); and if all the strategy amounts to is providing us with a physical reason why there an inferential disconnect, then it may achieve the former but will fail to establish the latter.[23] It is true that *if* phenomenal states turned out to be physical states, then zombies (for example) would be impossible, even though they are conceivable; but simply giving a physical account of the unusual nature of our phenomenal concepts provides no reason for believing in this identity nor does it give us any explanation as to *how* phenomenal states could be identical to physical states – hence it does nothing to support a rejection of Chalmers' argument against type-B materialism.

What might change the situation so far as Chalmers is concerned is if the materialist were to provide: (i) some *reason* for thinking that the two different types of state can somehow be identical, or (ii) an argument which shows that epistemic gaps regarding true beliefs *can* arise in a purely physical system – since that would be a clear counterexample to the thesis that the epistemic gaps must imply an ontological gap. And Chalmers' argument against the PCS is designed to show that it is impossible for that strategy to provide the latter.

(f) The case of interiority

We are now in a position to consider the relevance of Chalmers' critiques, of the PCS and of type-B materialism, to an explanation of the epistemic gaps in terms of interiority.

---

[23] That this is the requirement which Chalmers demands of the PCS can be clearly seen in his discussion of Hill & McLaughlin's suggestion (1999) that the epistemic gaps arise from two independent cognitive processes – one process by which we conceive of physical processes and another by which we conceive of phenomenal processes: '[Hill & McLaughlin's account] explains why zombies might be conceivable, *but it does nothing to explain why and how this conceivability coexists with the impossibility of zombies*. If we do not demand such an explanation, a strategy like Hill and McLaughlin's could be used to "explain away" any conceivability intuition at all. Presumably, there will always be a psychological explanation of the processes involved in a modal intuition, but one should not infer that these intuitions are always unreliable. If one did, then one might likewise find a psychological explanation of our mathematical beliefs and infer that these beliefs are no guide to mathematical truth.' (2010, p. 182, emphasis added)

In section 4, I argued that there is good reason to *expect* that a particular sort of epistemic gap might arise in a physical universe containing complex living organisms: namely, a gap which reflects the dynamic co-development of interiority and exteriority, and therefore of two different sorts of concept – those concepts which have first-person applicability criteria, and those with objective, third-person applicability criteria. However, my argument in section 4 is not intended as a proof that such a gap can arise in a world *which is describable in purely physical terms*, since it may well be that the gap itself is not describable without employing first-person terminology. This is similar to the point that I mentioned in respect of Chalmers' argument against the PCS: if we restrict ourselves to topic-neutral (i.e. non-phenomenal) terms, then doesn't this imply that my epistemic situation is the same as that of my zombie twin, and that *neither* situation incorporates any epistemic gap?

So the aim of my argument in section 4 is *not* to prove that an epistemic gap can exist in a wholly physical world – it is instead targeted at Chalmers' general reasons for rejecting type-B materialism, and in particular his dismissal of identity claims as unmotivated and unjustifiable. If my arguments in sections 3 and 4 are right, then the only epistemic gap in respect of consciousness is that which results from our having two different sorts of functionally characterized concepts: (i) the first-person concepts which we derive from our own experience of recognizing, discriminating, and so on; and (ii) the objective concepts of recognizing, discriminating, and so on, which are describable in purely third-person terms – for example, in terms of observable behaviour.[24]

But we *do* have a rationale for thinking that there is some form of identity underlying the properties which instantiate these two types of concept: namely, that the mental processes of recognizing, discriminating, etc. that we conceptualize in first-person terms are *the very same mental processes* of recognizing, discriminating, etc. that we can conceptualize in third-person terms. Of course, Chalmers can try and repeat his strategy of arguing that the epistemic intensions of the former concepts are different from the epistemic intensions of the latter, but this would now be beside the point. We accept that the epistemic gap exists, but have a clear rationale for maintaining that there is an identity nonetheless: the conceptual gap we are concerned with here is not that between, say, my concept of 'bluish' and my concepts of neuronal activity, but between two different concepts of recognizing, two different concepts of discriminating, and so on. For Chalmers to insist that the relationship between the properties that instantiate these two different types of concept must be that of a fundamental law

[24] My discussion in section 5 of the thought experiments which feature in *The Conscious Mind* suggests that, although zombies *are* conceivable, the other thought experiments fail to establish any epistemic gap – because Mary learns *no* new propositional fact, and inverted qualia are *not* conceivable. The latter two thought experiments depend in part upon an implicit assumption (which it is open to the materialist to reject) that there exist intrinsic phenomenal properties – i.e. properties which Mary can acquire knowledge of, and which can be inverted with respect to any functionally characterised property. The difference between the zombie thought experiment and the other two thought experiments arises because it is *the interior world itself* that cannot be entailed a priori by the physical facts, and any lack of entailment of the *specific qualitative aspects* of experience is derivative of this gap. So zombies, who entirely lack an interior world, are conceivable; but once we accept such an interior world, it is (I argued in section 5) *not* conceivable that the specific qualitative aspects of an experience can be left undetermined by that experience's functionally characterised properties.

of nature, rather than that the properties represent two different ways of apprehending a single mental process, would appear to be unmotivated – except by a desire to save his argument for dualism. It is now Chalmers who might be described as putting forward a position that is 'at best ad hoc and mysterious' – particularly when we recall that what is at issue is not the contrasting merits and demerits of materialism and dualism, but whether materialism is even *possible*.

(g) Objective and phenomenal senses of interiority

What was being argued for in section 4 is that we can give a physical account of interiority: i.e. of how it comes to be that some of the phenomena that present themselves to us are not capable of producing inter-subject agreement since they are cognitively available to only a single organism. There is nothing here, it will be noted, which relates to *phenomenal* states as such. For example, our explanation of interiority might be in terms of the evolutionary development of what Chalmers refers to as 'perceptual and introspective processes'.

Now as far as Chalmers is concerned, all mental processes are functionally definable in objective terms – and therefore even interiority can be defined in terms of behaviour, such as an organism acting in certain ways in response to what is going on in its body and nervous system. Of course, there is *something it is like* to have the interiority of a human being: e.g. something it is like to *perceive* a large object moving towards me, something it is like to *introspectively* consider my memory of what happened yesterday, and something it is like to *recognize* the colour I am experiencing as that which I call 'blue'. But these are all mental activities considered from the interior perspective, rather than from the objective perspective of observable behaviour, and I cannot a priori deduce interiority in this *phenomenal* sense from the physical facts alone. But the reason for this is the same as the reason why I cannot deduce the phenomenal facts from the physical facts in general – the applicability criteria for our concepts of the former are not capable of providing the inter-subject agreement by which facts are constituted as 'objective' facts. So when we consider interiority in this phenomenal sense, we find ourselves back in our original situation: we cannot form an a priori link between the two different types of description of a mental state or activity – one description that is internally characterized, and another description which is characterized in terms of externally observable behaviour. This is as true of interiority as it is of any other psychological feature.

Our stance should now be that taken by Balog (2012): *if* it is the case that dualism is wrong, and that phenomenal states are indeed physical or functional states of an organism, then we have plausible physical explanations as to why we are presented with the epistemic gap which the anti-materialist's arguments rely on. So that epistemic gap cannot lead to the ontological conclusion that it is *impossible* for phenomenal states to be physical or functional states. Chalmers' rejection of this type of argument – on the grounds that it invokes a primitive identity that is brute and unmotivated – has been undercut by the fact that the identity now invoked is based upon two different ways of apprehending a mental

activity. And this is an identity which, I have argued, *cannot* be simply dismissed as brute or unmotivated.

7: Interiority, exteriority, ontology

(a) The privacy of consciousness reconsidered

The discussion in section 4 centred on the apparent impossibility of inter-subject agreement on phenomenal facts, and therefore of objective applicability criteria for phenomenal concepts. The lack of such criteria would prevent there being any a priori entailment between (i) facts concerning mental processes described in terms of objective behaviour, and (ii) facts concerning those same processes considered as our own experienced mental activities. However, there are those who believe that some form of inter-subject agreement concerning interior facts cannot be ruled out. Keith Frankish, for example, suggests (2015) that *introspective subjectivity*, which he characterizes as 'an awareness of one's own mental responses to the world' and 'to know what you are like as you respond to the world', is not *essentially* private: '[w]ith the right apparatus, another person might monitor the same internal states my introspective mechanisms do and so share my introspective awareness'. One could extend this thought experiment to mechanisms for sharing perceptual stimulae and so on, such that the other person has exactly the same perceptions and introspective awareness as I do; and one might speculate that this could enable the other person to agree with me about facts concerning my interior mental processes. Does this sort of thought experiment enable us to conclude that inter-subject applicability criteria for phenomenal concepts might be possible after all? I am not convinced that it does, because we can always ask the question – is there one subject or two? If the former, then there is still no *inter-subject* agreement on phenomenal facts; if the latter, then each subject is applying their phenomenal concepts to their *own* experiences.

What this suggests is that the fundamental reason for the epistemic gap between the physical and the phenomenal lies not in the qualitative aspects of the latter, nor in the lack of third-person cognitive access to the experiences of an individual organism, but in the *existence* of interiority itself. Suppose, just for one moment, that some sort of inter-subject knowledge concerning my interior mental activities *were* somehow possible. It does seem that no objective, third-person criteria could ever overrule the first-person criteria that I have for applying my concepts of those interior processes. For example, my criteria for saying that I am recognizing the colour which I am now experiencing as that which I term 'blue' is simply that I am aware that that is precisely what I am doing. Why would I instead use some roundabout route that relied on objective criteria, such as those relating to the observation of my own behaviour and the scientific measurement of the activity of the neurons in my

brain?[25] Or – supposing that it were somehow possible – the judgments of others who have access to my interior mental states? There would still seem to be no route for any a priori entailment from objectively described facts to facts concerning my own mental processes as I experience myself performing them.

If this is right, then rather than racking our brains further on the issue of the entailment or non-entailment of the phenomenal facts, we might gain more insight if we were to view mental processes as being *essentially* first-person processes – a position which John Searle (1992, pp. 16-17) has always maintained:

> The way that the third-person point of view is applied in practice makes it difficult for us to see the difference between something really having a mind, such as a human being, and something behaving *as if* it had a mind, such as a computer. And once you have lost the distinction between a system's really having mental states and merely acting as if it had mental states, then you lose sight of an essential feature of the mental, namely that its ontology is essentially a first-person ontology.

It is for losing sight of this basic fact about consciousness that Searle (1997, pp. 160-161) criticizes Chalmers' assertion that there are distinct phenomenal and functional senses of the terms used to describe mental states:

> There are not two definitions of the psychological terms such as 'belief,' 'desire,' 'pain,' and 'love,' one definition referring to conscious states, one to material states. Rather, only systems capable of consciousness can have any psychology at all, and though all of us have plenty of unconscious mental states, unconscious beliefs and desires for example, we understand these as mental states, because we understand them as potentially conscious, as the sorts of thing that might have been conscious but are not because of repression, brain damage, or perhaps just because we fell asleep.

As an example, consider the mental function of 'attention'. My main knowledge of this function is first-person knowledge in the sense that I would have little understanding of what, say, an 'attentive' robot was actually doing unless I had experienced the way that I myself can focus or change my own attention. I might see that the robot was looking at one thing, and then looking at another thing, but I wouldn't be able to grasp the notion of 'attention' – the robot's behaviour would be mysterious to me. So in this sense, the first-person perspective of a mental function would seem to be the primary perspective. Evan Thompson (2007, p. 163) makes a similar point about the primacy of the first-person viewpoint, referring to an example used by Hans Jonas:

> Let us suppose, following Jonas (1966, pp. 64-98), that we were looking at an organism from the perspective of a disembodied and purely analytical, mathematical intellect. Would we be able to recognize the organism's inwardness and purposiveness? From the disembodied, analytical standpoint,

[25] Sometimes I do learn of my own mental states in this manner. For example, I may not realize that I am embarrassed until I see myself in a mirror and realize that my face is red. However, my concept for 'feeling embarrassed' is unlikely to include 'having a red face' amongst its applicability criteria – even if it might include 'my face feels flushed'.

> the organism would be resolved into a collection of fleeting, objective physicochemical events, 'and all features of a *self-related autonomous unity* would, in the end, appear as purely ... fictitious' (p. 78, emphasis added). Yet we are bodily beings ourselves, and we experience inwardness and purposiveness in our dealings with the world. Thus, '*we* are able to say what no disembodied onlooker would have a cause for saying: that the mathematical God in his homogeneous analytical view misses the decisive point—the point of life itself: its being self-centered individuality, being for itself and in contraposition to all the rest of the world, with an essential boundary dividing "inside" and "outside"' (p. 79).

(b) From epistemology to ontology

Thompson's distinction between 'inside' and 'outside' is not just a matter of epistemology – that is, of *understanding* or *knowing* from an interior or exterior perspective. The differentiation between interiority and exteriority is also of particular significance when examining the conclusion of Chalmers' argument against materialism: i.e. the falsity of materialism as an *ontological* thesis – to which I now turn.

Chalmers characterizes materialism and dualism as ontological propositions in two ways: firstly in terms of *facts* (although he occasionally refers to *truths* rather than facts), for example:

> In our language, materialism is true if all the positive facts about the world are globally logically supervenient on the physical facts. This captures the intuitive notion that if materialism is true, then once God fixed the physical facts about the world, all the facts were fixed. (1996, p. 41)

and secondly, in terms of *properties*:

> The dualism implied here is instead a kind of *property* dualism: conscious experience involves properties of an individual that are not entailed by the physical properties of that individual, although they may depend lawfully on those properties. (1996, p. 125)

The connection between the two is that the totality of all facts includes facts about the instantiation of properties. So if q is a phenomenal property, then there is a *fact* as to whether or not q is instantiated in someone's experience (or would be instantiated in a specified possible world).

Chalmers argues that the epistemological gaps between the phenomenal and the physical straightforwardly imply the falsity of materialism: i.e. that there are facts which are not completely determined by the physical facts (1996, pp. 123-124). Furthermore, if phenomenal properties are not determined by physical or functional properties, then we have to consider them to be *intrinsic* in nature – which implies a dualism of properties. However, Chalmers is uncertain as to whether there is also a dualism of 'substances'. In *The Conscious Mind* he claimed that 'the issue of what it would take to constitute a dualism of substances seems quite unclear to me' (1996, p. 125). But his later, more considered opinion is that substance dualism revolves around the question as to whether there are fundamental phenomenal *objects* or *individuals*, in addition to the objects or individuals of physics.

The alternative would be that there are two different types of *properties* which attach to a single type of fundamental object or individual. On this matter, Chalmers says, he is neutral (2010, p. 139).

Now I have suggested the ineliminable epistemic gap which we find in respect of consciousness reflects the fact that we have two different ways of thinking about a process such as ‘attention’ or ‘recognition’ – firstly as a mental activity which we experience ourselves performing, and secondly in terms of objectively defined, observable behaviour – and that the distinction between these two sorts of notion of a mental process is directly related to the development of interiority. But the evolutionary emergence of interiority is intimately connected with the emergence of *an exteriority* with respect to that interiority. An organism’s process of self-production regulates its interaction with the environment outside itself, i.e. the ways in which matter and energy flow through its boundary or membrane. An organism is therefore not just a system which internally self-produces itself – it is also an interactive agent in its environment. Hence, as Evan Thompson observes (2007, p. 79), the interiority displayed by even the simplest of organisms is:

> the self-production of an inside that also specifies an outside to which it is normatively related ... Thus the dynamic emergence of interiority can be more fully described as the dynamic co-emergence of interiority and exteriority.

In other words, there can be no ‘outside’ without an ‘inside’.

This implies that it is not just exteriority which is intimately linked with the emergence of interiority – *so is ontology itself*. Objects and properties as such – carved out, as it were, from the rest of the universe – cannot be thought of as existing without some form of interiority which distinguishes them. This is the view developed by the theoretical biologist Humberto Maturana, whose concept of autopoiesis forms one of the central pillars of Evan Thompson’s account of the co-emergence of interiority and exteriority (see Thompson 2007, chapters 5 and 6). For Maturana, a basic cognitive activity is that of *distinction*, whereby an organism separates out object from background. But precisely because distinctions are made by living entities, objects cannot be said to exist in a domain that is conceptually prior to, or independent of, the observer:

> ... without observers nothing exists, because existence is specified in the operation of distinction of the observer. For epistemological reasons, we ask for a substratum that could provide an independent ultimate justification or validation of distinguishability, but, for ontological reasons, such a substratum remains beyond our reach as observers ... [However,] once a domain of reality is brought forth, the observer can treat the objects or entities that constitute it both as if they were all that there is and as if they existed independently of the operations of distinction that bring them forth. (Maturana, 1988)

The account of the evolutionary development of interiority given by the psychologist Nicholas Humphrey is equally explicit on this point. Before the emergence of sentient beings, writes Humphrey (1992, pp. 16-17):

> ... the phenomena that we now call the phenomena of the material world were not yet in existence: ... because no one was there, there was not ... anything that *counted as* a volcano, or a dust-storm and so on. I am not suggesting that the world had no substance to it whatsoever. We might say, perhaps, that it consisted of 'worldstuff'. But the properties of this worldstuff had yet to be represented by a mind.

This position reflects a long train of philosophical thought, among whose recent exponents are neo-pragmatists such as Richard Rorty and Hilary Putnam. Rorty's book *Philosophy and the Mirror of Nature* (1979) is a sustained attack on the coherence of the idea that knowledge consists of *accurately mirroring* in our minds the world outside, an achievement that is possible only through those special mental processes that supposedly form the backbone of Western philosophy and of the methodology of modern science. According to Rorty (1979, p. 12), our philosophy has been unduly dominated by Greek ocular metaphors:

> The picture which holds traditional philosophy captive is that of the mind as a great mirror, containing various representations – some accurate, some not – and capable of being studied by pure, nonempirical methods. Without the notion of the mind as mirror, the notion of knowledge as accuracy of representation would not have suggested itself. Without this latter notion, the strategy common to Descartes and Kant – getting more accurate representations by inspecting, repairing, and polishing the mirror, so to speak – would not have made sense.

Giving up the notion of the mind acting as 'the mirror of nature' implies two theses argued for by Hilary Putnam (1981; 1987; 1988; 1999): firstly, that there is no single correct description of reality, and secondly that our objectifications of things as individual entities, separated out from the rest of reality, depend on our cognitive processes, which in turn arise from our interactions with the world.[26]

The notion that there is a single correct description of reality is the product of two things: (i) our naive realism concerning the world about us, transmuted into a metaphysical thesis; and (ii) the dualism that has permeated Western thought since Descartes, with two domains (variously characterized as matter/objective/noumenal on the one hand, and as soul/subjective/phenomenal on the other) but little in the way of a coherent method of linking the two other than the use of phrases to the effect that the latter domain 'reflects', or 'represents', or 'corresponds to' the former domain. The lack of a unique description of reality leads Putnam to conclude that the question *what objects does the world consist of?* can only make sense *within* a specified description or theory; hence, there can be no 'God's Eye View' which can provide us with *the* truth about the way the world is (Putnam, 1981, pp. 49-50). This is not a denial of realism *tout court*; Putnam emphasizes that there *is* a reality separate from our own thoughts, but maintains that there is no single, objectively correct, description of that reality from an external point of view.

[26] I explore Putman's arguments in 'Mirroring Reality: How *Not* to Think about the World' (Holliday, 2005).

This idea that the world is not simply divided up into definite objects which then make themselves available to our perceptual and cognitive processes, but rather that our identification of things as entities is itself dependent upon those processes and our interactions with the world, is widely shared by researchers working in a variety of fields. For example:

a) The phenomenologist Maria Villela-Petit (1999, p. 513):

> A stone or a galaxy does not open any world. It is not for a stone that there is another stone; it is only for a living being that there can be stones or any such objects ...

b) Stan Franklin (1995, p. 300), an artificial intelligence researcher:

> Things don't come with bar codes so that I can tell what I'm sitting on is a chair. Neither do categories such as water, clouds, calculus class, quarks.

c) Professor of neurology and Zen adept James Austin (1998, p. 550):

> We and the pebble are still integral parts of the same ongoing universe. *It is still a universe undivided.* It has never been split up by such artificial distinctions as minutes, first names, rock names, or place names. It *is*, in its suchness.

If this line of thought is right, then our metaphysics must not take for granted the category of 'object', and thereby presume *at the outset* a distinction between 'subject' and 'object'. This would be a clear case of what Brian Cantwell Smith (1996, p. 50) terms an 'an inscription error', whereby we make – explicitly or implicitly – an initial ontological assumption, and then later read back that assumption or its consequences as constituting an empirical discovery or theoretical conclusion, without recognizing the dependence on the original assumption. We would have presupposed the existence of a realm of exteriority together with an interior realm which in some way refers to, or reflects, that exterior realm; and by doing so, we would have taken for granted within our fundamental metaphysics the existence of reference and other forms of intentional participation in the world, and so be unable to explain them.

This, I suggest, is what Chalmers does by demanding that an explanation for the phenomena of interiority be given wholly in terms of the phenomena of a *taken-for-granted exteriority*, rather than in terms of *the development of interiority-exteriority as a form of dialectical unity*. We saw in section 4 that the reason for the epistemic gap between physical descriptions and phenomenal descriptions is that we have two different ways of thinking about a *single* psychological process, such as 'recognition'; firstly as a mental activity which we ourselves perform, and secondly in terms of objectively-defined, observable behaviour. And *this* distinction is related to the co-emergence of interiority and exteriority: it is not as if exteriority – and therefore 'recognition' as an observable behaviour – already exists, and we have to account for how interiority – and 'recognition' as one of our own mental processes – arises *from* that exteriority. Instead, we are using our existing language of

exteriority as a 'lodging-place'[27] in order to understand the emergence of both interiority and its dialectical counterpart, exteriority – and thereby understand how a mental process such as 'recognition' can be apprehended in two different ways. The fact that in doing so we utilize the language of exteriority does *not* imply that we thereby accept the existence of an external realm of objects and properties which is ontologically prior to a realm of interior phenomena.

8: Conclusion

Chalmers' case against materialism is based around the epistemic gap which is supposed to exist between phenomenal facts and physical or functional facts, a gap which is revealed in different ways, depending on the precise argument deployed – sometimes it manifests itself in terms of explanation or analysis, sometimes in terms of conceivability, sometimes in terms of our knowledge of the facts. However, by carrying out an 'archaeology' of Chalmers' arguments, we have brought to light a number of singular features of this putative gap.

Firstly, we discovered that Chalmers fails to properly distinguish between our intuitive characterization of the qualitative aspect of an experience and the applicability criteria for our concept of that qualitative aspect. This failure makes it appear plausible that phenomenal concepts differ so completely from physical concepts that there can be no possibility of any 'conceptual hook' between the two types of concept, and hence no a priori entailment from physical or functional facts to phenomenal facts. However, Chalmers presents no argument to show that the applicability criteria for phenomenal concepts cannot be expressed in functionally characterized terms; indeed, it is both possible and plausible that the applicability criteria for, say, my use of the concept 'bluish' are expressible in terms of my recognition of the colour I am experiencing as that which I term 'blue' and my discrimination of that colour's relationship to other colours which I experience. What this implies is that, contrary to first impressions and to the way in which Chalmers usually frames his argument, the qualitative aspects of consciousness cannot of themselves generate an unbridgeable epistemic gap between phenomenal facts and physical or functional facts.

Secondly, and closely related to the first feature, is Chalmers' tendency to implicitly assume that phenomenal properties must be *intrinsic* properties. There is some debate about just what it is that makes a property 'intrinsic' – see, for example, the discussion by Weatherson & Marshall (2012). However, the basic notion is well-described by Yablo (1999): 'You know what an intrinsic property is: it's a property that a thing has (or lacks) regardless of what may be going on outside of itself'. If we combine this characterization with Chalmers' definition of materialism as the thesis that *all* facts are fully determined by the physical facts, then the existence of intrinsic phenomenal properties would

[27] The lodging-place metaphor is taken from Chuang Tzu: 'Saying from a lodging-place ... You borrow a standpoint outside in order to sort a matter out' (Graham, 1989, p. 106).

appear to *ipso facto* rule out materialism. This means that any argument against materialism which assumes that phenomenal properties are intrinsic properties is open to the objection that it simply begs the question. This is presumably why Chalmers carefully defines qualia as being simply those properties which type experiences by what it is like to have them, and it is the reason why I have used the phrase ‘a bluish aspect of my experience’: both are intended to avoid any question-begging assumption as to the nature of the qualitative properties of experiences. However, thinking about phenomenal properties in terms of their intuitive characterizations, rather than in terms of the applicability criteria of their concepts, makes it easy to view them, perhaps unwittingly, as intrinsic properties. Unfortunately, in making his case against materialism Chalmers appears to do precisely this at a number of key points: (i) in his neglect of any systematic investigation of the applicability criteria for our concepts of phenomenal properties (he simply takes the criterion to be the instantiation of the relevant intrinsic phenomenal property – as discussed in section 3); (ii) when he has recourse to inverted qualia (without intrinsic properties there is nothing to be inverted – as explained in section 5); and (iii) in his use of the knowledge argument (the new knowledge that Mary supposedly learns is knowledge about *which* intrinsic phenomenal property is instantiated when other people observe, say, a red rose – as also noted in section 5).

Thirdly, there cannot be objective, third-person criteria for the use of concepts relating to phenomena which by their very nature are only cognitively available to one person. The privacy of consciousness means that the only person who is in a position to say whether their concept of ‘bluish’ applies to *my* experience is myself. Hence, concepts of phenomenal qualities cannot have objective applicability criteria, i.e. criteria which might be evaluated by any competent and suitably-situated third party, and it is *this* consideration which results in an ineliminable epistemic gap between the phenomenal and the physical or functional. However, the private nature of consciousness is scarcely mysterious, and can be given a plausible explanation in terms of the evolutionary development of interiority as a feature of complex organisms.

Fourthly, Chalmers assumes that any materialist explanation of consciousness must be provided in terms of an already-existing assemblage of objects and properties, rather than in terms of the co-development of interiority and exteriority as a dialectical unity. This takes it for granted that the external realm is, ontologically-speaking, prior to the interior realm: an unwarranted assumption, especially when applied to mental activities that can be conceptualized from *both* interior and exterior perspectives.

These four factors all contribute to the epistemic gap as Chalmers sees it, and their intertwining creates a significant difficulty for anyone who hopes to make a coherent and focussed rebuttal of Chalmers’ position. This problem is exacerbated by the fact that, at one time or another, Chalmers has put forward several different arguments against materialism: one might address one aspect, or one particular version, of his argument, but then be subject to objections based upon other aspects or other

versions. The same applies to anyone attempting to provide a materialist account of consciousness which evades Chalmers criticism. Two examples of this – both of which I have referred to above – are Austen Clark's account of phenomenal properties in terms of quality spaces (Chalmers: 'this method does not explain the *intrinsic* nature of a color experience' – 1996, p. 235) and Nicholas Humphrey's account of the evolutionary development of interiority and the privacy of consciousness (Chalmers: '[t]he relevant step in the explanation is typically passed over quickly, however, and usually ends up looking something like magic. After some details about information processing are given, experience suddenly enters the picture, but it is left obscure *how* these processes should suddenly give rise to experience' – 2010, p. 12). Hence the importance of separating out the various strands and versions of Chalmers' case against materialism and dealing with each separately, utilising accounts of quality space, the evolution of interiority, and so on, in the appropriate place. In my own critique of Chalmers, for example, quality spaces contribute to an understanding of the applicability criteria for phenomenal concepts, and the evolution of interiority contributes to an understanding of how we can have two different sorts of functionally characterized concepts of mental activities, with no a priori entailment between them. Such an approach can utilize the insights of the likes of Austen Clark or Nicholas Humphrey, without viewing them as providing a complete explanation of experience.

Chalmers himself does not separate out the different strands which contribute to the epistemic gap as he conceives of it. As a result, he is too easily convinced that there is a complete conceptual discontinuity or incongruence between the qualitative aspects of experiences and physical or functional properties, and concludes that it must therefore be impossible to gain any explanation of the former by means of a consideration of the latter. However, once we disentangle the separate strands, our perspective is transformed. It is now unconvincing to maintain, as does Chalmers (1996, p. xi), that it is *utterly mysterious* as to how a biological organism could come to have conscious experiences. To maintain this position on the basis of a lack of a priori entailment between (i) psychological processes describable in terms of objective behaviour, and (ii) those same processes describable as one's own experienced mental activities, would appear simply obtuse – unless one were *already* convinced of the truth of dualism or panpsychicism.

What this 'archaeology' points towards is the conclusion that the epistemic gap described by Chalmers has more to do with the nature of our own epistemological and ontologizing activities than it does with 'the basic furniture of the world' (2010, p. 18). In which case, a fuller account of consciousness will require a better understanding of those enterprises, rather than an acceptance of some form of dualist or panpsychic ontology.[28]

[28] Or as Güven Güzeldere (1997, p. 45) put it, nearly twenty years ago: 'What seems the most promising direction in reapproaching consciousness and pursuing its deep-rooted problems in the present era involves rethinking epistemology and conceptual schemes (as opposed to a priori postulation of new ontology) to yield a cross-fertilization of the first-person and third-person perspectives ...'

References

Austin, J.H. (1998) *Zen and the Brain*, Cambridge, MA: MIT Press.

Alward, P. (2004) Is phenomenal pain the primary intension of 'pain'?, *Metaphysica,* **5** (1), pp. 15-28.

Balog, K. (2012) In defense of the phenomenal concept strategy, *Philosophy and Phenomenological Research* **84** (1), pp. 1-23.

Block, N. (2006) Max Black's objection to mind-body identity, *Oxford Studies in Metaphysics*, **2**, pp. 3-78.

Block, N. (2015) The Canberra Plan neglects ground, in T. Horgan, M. Sabates & D. Sosa (eds.) *Qualia and Mental Causation in a Physical World: Themes from the Philosophy of Jaegwon Kim*, Cambridge University Press.

Block, N. & Stalnaker, R. (1999) Conceptual analysis, dualism, and the explanatory gap, *Philosophical Review*, **108** (1), pp. 1-46.

Byrne A. (2010) Inverted qualia, in Zalta E.N. (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2010 edition), [Online], http://plato.stanford.edu/archives/spr2010/entries/qualia-inverted/ [8 Dec 2013].

Carruthers, G. & Schier, E. (2014) Why are we still being hornswoggled? Dissolving the hard problem of consciousness, *Topoi*, Advance online publication: DOI http://link.springer.com/article/10.1007%2Fs11245-014-9266-3.

Carruthers, P. (2004) Reductive explanation and the 'explanatory gap', *Canadian Journal of Philosophy*, **34** (2), pp. 153-174.

Carruthers, P. & Veillet, B. (2007) The phenomenal concept strategy, *Journal of Consciousness Studies*, **14** (9-10), pp. 212-236.

Chalmers, D.J. (1995) Facing up to the problem of consciousness, *Journal of Consciousness Studies*, **2** (3), pp. 200-219.

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

Chalmers, D.J. (1997) Moving forward on the problem of consciousness, *Journal of Consciousness Studies*, **4** (1), pp. 3-46.

Chalmers, D.J. (1999) *Philosophy 596B: mind and modality (spring 1999), week 2 discussion*, [Online], http://consc.net/class/596b/week2.txt [21Oct 2014].

Chalmers, D.J. (2002a) On sense and intension, *Philosophical Perspectives*, **16** (s16), pp. 135-182.

Chalmers, D.J. (2002b) Does conceivability entail possibility?, in T.S. Gendler & J. Hawthorne (eds.) *Conceivability and Possibility*, Oxford University Press.

Chalmers, D.J. (2003) The content and epistemology of phenomenal belief, in Q. Smith & A. Jokic (eds.) *Consciousness: New Philosophical Perspectives*, Oxford University Press.

Chalmers, D.J. (2004) Response to Perry's reply, [Online], http://consc.net/papers/perry.html#response [21 Oct 2014].

Chalmers, D.J. (2006) The foundations of two-dimensional semantics, in M. Garcia-Carpintero & J. Macia (eds.) *Two-Dimensional Semantics*, Oxford University Press.

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

Chalmers, D.J. (2012a) Comment on 'dissolving the hard problem of consciousness', *Consciousness Online fourth conference*, [Online], http://consciousnessonline.com/2012/02/17/dissolving-the-hard-problem-of-consciousness/#comment-1365 [21 Oct 2014].

Chalmers, D.J. (2012b) *Constructing the World*, Oxford University Press.

Chalmers, D.J. & Jackson, F. (2001) Conceptual analysis and reductive explanation, *Philosophical Review*, **110** (3), pp. 315-361.

Churchland, P.M. (1996) The rediscovery of light, *Journal of Philosophy*, **93** (5), pp. 211-228.

Clark, A. (1993) *Sensory Qualities*, Oxford University Press.

Clark, T.W. (2005) Killing the observer, *Journal of Consciousness Studies*, **12** (4-5), pp.38-59.

Clark, T.W. (2010) Respecting privacy: why consciousness isn't even epiphenomenal, [Online], http://www.naturalism.org/philosophy/consciousness/respecting-privacy [26 November 2015]

Damasio, A. (2010) *Self Comes to Mind: Constructing the Conscious Brain*, London: William Heinemann.

Deacon, T.W. (2012) *Incomplete Nature: How Mind Emerged from Matter*, New York: W.W. Norton & Company.

Dennett, D.C. (1991) *Consciousness Explained*, Boston, MA: Little, Brown.

Dennett D.C. (1993) The message is: there is no medium, *Philosophy and Phenomenological Research*, **53** (4), pp. 919-931.

Dennett, D.C. (2005) *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*, Cambridge, MA: MIT Press.

Dennett, D.C. (2012) Comments at the *Moving Naturalism Forward* conference, [Online]. Available on a video of the conference's first session at https://www.youtube.com/watch?v=Ju4C_ITlBsU [25 March 2016].

Diaz-Leon, E. (2010) Can phenomenal concepts explain the epistemic gap?, *Mind*, **119** (476), pp. 933-951.

Doyle, A.C. (1981) *The Penguin Complete Sherlock Holmes*, Harmondsworth, Middlesex: Penguin Books.

Elpidorou, A. (2013) Having it both ways: consciousness, unique not otherworldly, *Philosophia*, **41** (4), pp. 1181-1203.

Frankish. K. (2015) Something that it is like to be, *Tricks of the Mind*, [Online], http://www.keithfrankish.com/2015/04/something-that-it-is-like-to-be/ [14 April 2015]

Franklin, S. (1995) *Artificial Minds*, Cambridge, MA: MIT Press.

Goff, P. (2011) A posteriori physicalists get our phenomenal concepts wrong, *Australasian Journal of Philosophy*, **89** (2), pp. 191-209.

Graham, A.C. (1989) *Chuang-Tzu: The Inner Chapters*, London: Unwin-Hyman.

Güzeldere, G. (1997) The many faces of consciousness: A field guide, in N. Block, O. Flanagan & G. Güzeldere (eds.) *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press.

Hardin, C.L. (1988) *Color for Philosophers: Unweaving the Rainbow*, Indianapolis: Hackett.

Henderson, R.W. (2014) Breaking the spell: materialism and the qualia intuition, *Journal of Consciousness Studies*, **21** (7-8), pp. 184-192.

Hilbert D.R. & Kalderon M.E. (2000) Color and the inverted spectrum, in S. Davis (ed.) *Color perception: philosophical, psychological, artistic, and computational perspectives*, Oxford University Press.

Hill, C.S. & McLaughlin, B.P. (1999) There are fewer things in reality than are dreamt of in Chalmers's philosophy, *Philosophy and Phenomenological Research*, **59** (2), pp. 445-454.

Holliday, M. (2005) Mirroring reality: how not to think about the world, [Online], http://www.holli.co.uk/metaphysics.htm [28 April 2015]

Humphrey, N. (1992) *A History of the Mind*, London: Vintage, paperback edition 1993.

Hutchinson, D. (2013) The origin of phenomenal consciousness: on the art of the 'hard problem', *Journal of Consciousness Studies*, **20** (1-2), pp. 126-152.

Jackendoff, R. (2012) *A User's Guide to Thought and Meaning*, Oxford University Press.

Jackson, F. (1982) Epiphenomenal qualia, *Philosophical Quarterly*, **32** (4), pp. 127-136.

Jackson, F. (1986) What Mary didn't know, *Journal of Philosophy*, **83** (5), pp. 291–295.

Jonas, H. (1966) *The Phenomenon of Life: Toward a Philosophical Biology*, Evanston, IL: Northwestern University Press, reprinted 2001.

Kriegel, U. (2009) *Subjective Consciousness: A Self-Representational Theory*, Oxford University Press.

Levine, J. (2001) *Purple Haze: The Puzzle of Consciousness*, Oxford University Press.

Lewis, D. (1988) What experience teaches, *Proceedings of the Russellian Society*, **13** (1), pp. 29-57.

Loar, B. (1997) Phenomenal states, in N. Block, O. Flanagan & G. Güzeldere (eds.) *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press.

Maturana, H.R. (1988) Ontology of observing: The biological foundations of self consciousness and the physical domain of existence, in *Texts in Cybernetics* (conference workbook), American Society For Cybernetics Conference, Felton, CA. 18-23 October 1988; available online at http://www.inteco.cl/articulos/004/texto_ing.htm [27 March 2016]

Nida-Rümelin, M. (1995) What Mary couldn't know: belief about phenomenal states, in T. Metzinger (ed.), *Conscious Experience*, Exeter, UK: Imprint Academic.

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

Papineau, D. (2006) Comments on Galen Strawson, *Journal of Consciousness Studies*, **13** (10-11), pp. 100-109.

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

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

Perry, J. (2004) Replies, *Philosophy and Phenomenological Research*, **68** (1), pp. 207-229.

Putnam, H. (1981) *Reason, Truth and History*, Cambridge University Press.

Putnam, H. (1987) *The Many Faces of Realism*, La Salle, IL: Open Court.

Putnam, H. (1988) *Representation and Reality*, Cambridge, MA: MIT Press.

Putnam, H. (1999), *The Threefold Cord: Mind, Body, and World*, New York: Columbia University Press.

Rabin, G. (2011) Conceptual mastery and the knowledge argument, *Philosophical Studies*, **154** (1), pp. 125-147.

Rorty, R. (1979) *Philosophy and the Mirror of Nature*, Princeton, NJ: Princeton University Press.

Searle, J.R. (1992) *The Rediscovery of the Mind*, Cambridge, MA: MIT Press.

Searle, J.R. (1997) *The Mystery of Consciousness*, London: Granta Books.

Smith, B.C. (1996) *On the Origin of Objects*, Cambridge, MA: MIT Press.

Thompson, E. (2007) *Mind in Life: Biology, Phenomenology, and the Sciences of the Mind*, Cambridge, MA: Harvard University Press.

Varela, F.J. (1997) Patterns of life: intertwining identity and cognition, *Brain and Cognition*, **34** (1), pp. 72-87.

Villela-Petit, M. (1999) Cognitive psychology and the transcendental theory of knowledge, in J. Petitot, F. Varela, B. Pachoud & J-M. Roy (eds.) *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, Stanford, CA: Stanford University Press.

Weatherson, B. & Marshall, D. (2012) Intrinsic vs. extrinsic properties, in Zalta E.N. (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), [Online], http://plato.stanford.edu/archives/fall2014/entries/intrinsic-extrinsic/ [19 Aug 2015].

Yablo, S. (1999) Intrinsicness, *Philosophical Topics*, **26**, pp. 479-505.