THE UNIVERSITY OF CALGARY

Phenomenal Properties: The Epistemology and Metaphysics of Qualia

by

Andrew R. Bailey

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

CALGARY, ALBERTA

JUNE, 1998

© Andrew R. Bailey 1998
THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a dissertation entitled “Phenomenal Properties: The Epistemology and Metaphysics of Qualia” submitted by Andrew R. Bailey in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Professor John A. Baker, Department of Philosophy (Supervisor)

Professor J.J. Macintosh, Department of Philosophy

Professor Robert X. Ware, Department of Philosophy

Professor G.C. (Cam) Tesky, Department of Psychology

Professor William E. Seager, Philosophy, University of Toronto

Date
Abstract

This dissertation develops and defends a detailed realist, internalist account of qualia which is consistent with physicalism and which does not resurrect the epistemological ‘myth of the Given.’ In doing so it stakes out a position in the sparsely populated middle ground between the two major opposing factions on the problem of phenomenal consciousness: between those who think we have a priori reasons to believe that qualia are irreducible to the physical (e.g. Nagel, Chalmers, Jackson), and those who implicitly or explicitly treat qualia as contentful but non-phenomenal physical properties (e.g. Dretske, Lycan, Armstrong, Tye).

I present a minimal, non-question-begging definition of “qualia” and then use this definition in a reformulation of the argument from perceiver relativity which shows that qualia must, at least in humans, be properties of states of the central nervous system. That is, brain states do not simply indicate the colours (for example) of external objects—they instantiate phenomenal colours. Since brain states are not coloured in the same way as are external objects, I argue that to token phenomenal property $F$ must be to be the first-person phenomenal sensation of $F$. I build on this position to argue that the phenomenal apprehension of qualia is “given”—immediate, certain and private—but that qualia nevertheless do not provide an absolutely certain epistemological foundation since they do not constitute a set of indisputable propositions about facts. I then define what it would be for physicalism to be true of qualia, but argue that the main a priori arguments for and against qualia physicalism—including the logical possibility of zombies and the impossibility of epiphenomenalism—fail. Whether qualia are consistent with physicalism, I claim, is still an open question, answerable only when we discover in more detail how qualia depend upon the brain.

Although strongly in sympathy with many commonsensical intuitions about qualia, this account stands in contrast to most influential extant theories of qualia, such as eliminativism (Dennett), externalism (Dretske), or property-dualism (Chalmers): I conclude the dissertation by using its results to argue explicitly against these three positions.
Acknowledgements

I am grateful for perceptive comments on (ancestors of) sections of this dissertation from Patricia Blanchette, Bob Bright, David Chalmers, Tim Crane, Eric Dayton, Fred Dretske, Marc Ereshefsky, Phil Hanson, Terry Horgan, Zoltan Jakab, Pete Mandik, Charlie Martin, Philip Pettit and Pete Zachar. Carole, my wife, has been my greatest support in every possible way. Many thanks, of course, are due my supervisor John Baker and to Jack Macintosh, Bill Seager, Cam Tesky and Bob Ware. Those errors that remain are entirely my own.
Table of Contents

Abstract................................................................................................................................. iii

Acknowledgements ............................................................................................................... iv

Table of Contents.................................................................................................................. v

Chapter 1: Introduction........................................................................................................ 1
  1. Mapping the Problem Space ............................................................................................ 4
  2. Outline of the Dissertation .............................................................................................. 13
  3. Terminology Matters ....................................................................................................... 15

Chapter 2: Arguments for ‘Sense-Data’ ............................................................................. 25
  1. The Sense-Datum Theory ................................................................................................. 25
  2. Arguments from Givenness .............................................................................................. 30
  3. Arguments from the Causal Nature of Perception ............................................................. 33

Chapter 3: Qualia and the Argument from Perceiver Relativity ....................................... 43
  1. The Argument from Illusion ............................................................................................ 44
  2. Objections to the Argument ............................................................................................ 49
     a) Premise ii): External Objects Have Incompatible Phenomenal Properties ................. 50
     b) Premise v): Phenomenal Properties are Instantiated in Misperception ..................... 53
     c) Steps ix) and x): The Inductive Generalisation ......................................................... 62
  3. The Argument from Hallucination .................................................................................... 63
  4. The Adverbial Critique .................................................................................................... 66
  5. Conclusion: Against ‘Figment’? ....................................................................................... 73

Appendix: Proof That Qualia Are Internal Properties ......................................................... 82

Chapter 4: The Immediacy of Qualia Apprehension ............................................................ 85
  1. No Intermediate Entity .................................................................................................... 91
  2. Being Uninterpreted ....................................................................................................... 94
  3. Being Uninferred ............................................................................................................. 102
     a) The Regress Argument .............................................................................................. 105
     b) Epistemological Analysis ......................................................................................... 106
     c) Certainty ................................................................................................................... 108
  4. Conclusion ...................................................................................................................... 110

Chapter 5: The Certainty of Qualia Apprehension ............................................................... 111
  1. Indubitability ............................................................................................................... 113
  2. Evidence ....................................................................................................................... 124
Chapter 1: Introduction

The philosophical term-of-art “ quale” is not a particularly clear or concrete one. To a first approximation, we can safely say that qualia are in some sense phenomenal properties; they are the felt or phenomenal qualities associated with conscious experiences such as the viewing of a colour, the hearing of a sound, the feeling of a pain, or the sensation of coldness. Imagine coming in on a hot day after a long run, pouring a cold glass of milk, and then drinking it: the whiteness of the milk, its coldness to the touch, the sound it makes being poured and its taste are all paradigmatic examples of qualia.

Up to this point most are in cautious agreement; the terrain beyond, however, is marked by widespread dissent. The core dispute is probably that between those who consider qualia to be a “Hard Problem” and those who do not. The Hard Problem camp are willing to talk of the phenomenal content of qualia; they hold that some mental states—those characterised by qualia, often called states of “phenomenal consciousness”—have more than just informational content: they also have a certain feel. For example, they argue, a sensation of redness is more than just a mental state that carries the information that something in the external world (the intentional object) is red; it is in addition a mental state that, so to speak, feels phenomenally red. The

---

1 I say “cautious,” since even the minimalist description of the previous paragraph can be (mis)construed in such a way that it becomes controversial. For example, we have defined qualia (in a preliminary way) as properties considered phenomenologically—considered “as they appear to” some observer. However this is not (or at least not yet) to say that qualia are phenomenal properties rather than being properties of the actual objects involved: we do not commit ourselves, by the very use of the term “qualia,” to the thesis that whiteness, for example, is an aspect of the perception of milk rather than being a property of the milk itself. This would be a controversial claim—one which could not be settled merely by stipulation, but for which a philosophical argument would be required. (As it happens, I shall be giving just such an argument for this conclusion in Chapter Three.)

redness is not just indicated by the mental state—it is, in some sense, tokened by that state; qualia are as much “show” as they are “tell.”

Tied up with this notion of a “raw feel,” for the typical exponent of the Hard Problem, are a bundle of attributes that frustrate the assimilation of qualia into the current physicalist worldview. Thus qualia are said to be ‘the last remaining Hard Problem’ for the physical sciences because they are thought to be essentially apprehensible only from the first-person perspective; because phenomenal content is necessarily incorrigible, private, immediate and ineffable; because we can conceive of the correlations between qualia and physical or functional states breaking down … but most of all just because qualia are phenomenal—they feel a certain way—and the remaining furniture of the universe (apparently) is not. The deep mystery for Hard Problem theorists, then, is something like the following:

How does conscious experience arise from a system of 100 billion well-connected neurons set in a soupy mixture of neurochemicals in which the only things that happen are activations of sets of neurons at different rates and changes in the composition of the soup? … Minimally, we want … a plausible picture of how a rambunctious bundle of nervous activity, set in its complex biochemical environment, can ever add up to experiences that are clearly individuated by the subjects who experience them as, for example, the taste of lemonade, the smell of a rose, the feeling of sadness and elation, and so on. (Flanagan 1992, 51)

On the other side, those who deny that qualia are a Hard Problem typically do not assert that they are an easy problem but they do hold that the mystery described by Owen Flanagan can ultimately be resolved within something rather like current-day physics: that a story can be told about neurones and chemicals (at the level, perhaps, of their information-carrying capacities) that will fully explain how we experience the feeling of pain and the taste of milk. Often, though not always, this confidence arises from the conviction that in fact there is no phenomenal content over and above what we have called informational content. Explain how certain mental states have the content that external objects are coloured, smelly, and tasty in certain determinate ways, and you have explained qualia. Show in sufficient detail how distributed areas of the brain ‘light up’ in response to certain stimuli, and how they bring about the consequent brain activity and efferent signalling, and you have said all that needs to be said about the sub-
jective taste of lemonade.

This dissertation is an effort to develop a model of qualia which takes very seriously their phenomenal nature, and the core intuitions behind talk of the Hard Problem, but which nevertheless holds well open the possibility that qualia, complete with their phenomenal content, can be integrated into a future physics—that the very same properties we apprehend as phenomenal redness or painfulness or smell can be described and explained from an ‘objective,’ third-person viewpoint. Though I certainly (and predictably) do not close the “explanatory gap” between qualia and the physico-chemical properties of the brain here, I do try to do two things towards bridging the chasm—or rather, towards putting in place some foundation materials on either side of the canyon, upon which a bridge might later be constructed. First, I attempt to build up a model of qualia which is in accord with widely and deeply held intuitions about qualia, and yet which does not possess what we might call the peripheral mystery-making properties of ontological ghostliness, special immediacy, epistemological certainty, necessary privacy, in principle irreducibility, and so on. Second, I try to develop an account of qualia that is sufficiently concrete, detailed, and maybe even plausible, that it could serve to guide future empirical work on the phenomenal properties of the brain, since it is on this front, if on any, that physico-chemical theories of qualia must eventually be hammered out. That is, I attempt to set

---

3 This phrase was introduced by Joseph Levine (1983) to label the putative fact that nothing we know or could conceive of knowing about the physical world can explain why qualia feel the way they do—that is, that there is no conceptual reduction of qualia currently available. The gap would be closed, then, by a conceptual reduction: by a story in, say, the language of physics which explained (which was also evidently a story about) the phenomenality of qualia.

4 “…[T]he central question is whether it makes sense to look for a method of conceiving and describing the features of consciousness that both fits into a common theoretical structure with neurophysiological concepts and is true to their nature, rather than being a travesty or denial of it: more grandly, can we discover or create a common vantage point from which it would be possible to see how subjective experience and neurophysiology are internally connected? We would have to begin from the prescientific way of thinking about consciousness and hope that further developments could be guided partly by knowledge of the physical conditions which indicate how the features of consciousness are organized and what their important causal connections are.” (Nagel 1993, 4–5). This is my project.
up the outlines of a target explanandum for the brain sciences, one which is not defined as in principle beyond their grasp, yet which does not ignore the full-blooded phenomenality central to qualia. I want to show that the epistemological and metaphysical model of qualia presented here is as plausible as any other, and is supportive of fruitful work.\(^5\)

The model of qualia I develop is mostly a positive, rather than a negative or reactive, account. However, it would be as well to begin by locating my project within the contemporary debate about qualia.

1. MAPPING THE PROBLEM SPACE

A useful way of dissecting the problem space for qualia is to ask the following sequence of diagnostic questions. First, given that qualia are defined (more or less) as phenomenal properties, \textit{of what} are they properties? There are four central kinds of candidate for what qualia are properties of:

a) Phenomenal individuals (such as sense-data, or souls);

b) Physical states of brain-like systems (e.g. particular sorts of neuronal activation, or higher-level information-processing states);

c) Physical individuals external to the brain (like trees and glasses of milk);

d) Nothing at all (i.e. qualia do not actually exist).

Of these options, b) seems by far the most popular among contemporary theorists (at least if this list is reasonably exhaustive,\(^6\) and if they are pressed to make a choice). Option a) is sub-

\(^5\) As Jerry Fodor once said, “The form of a philosophical theory, often enough, is: \textit{Let’s try looking over here}” (1981a, 31).

\(^6\) It is true that individuals which are neither physical nor phenomenal can be said to have properties—abstract individuals (like numbers), centres of gravity, or inexistent intentional objects, for example, might all be said to be property-holders which are neither physical nor phenomenal. However none of them seem to be plausible candidates for \textit{qualia}-holders—for being the individuals which are phenomenally white, taste phenomenally of milk, feel phenomenally cold, and so on. (Gilbert Harman (1990) defends the view that qualia are literally properties of intentional objects: just as what Ponce de León was searching for is the intentional object \textit{the Fountain of Youth}, what is green in visual experience is the intentional object \textit{the tree}. This is not quite externalism (option c) above)

This dissertation defends a particular form of option b), whereby qualia are properties of physical states of brain-like systems. I explicitly attack option c), externalism, in Chapter Three (I consider Dretske’s externalist arguments in detail in section two of Chapter Eleven), and also briefly consider there, and reject, option a), substance dualism. I implicitly contradict the eliminativist option d) throughout by developing a positive account of qualia which is, I hope, because, although for Harman it is the actual tree which is green when the tree exists, there need be no actual tree identical with the intentional object tree; as Harman admits, it could be a hallucination or a dream. Just what is the ontological status of these coloured, smelly, noisy, tasty intentional objects when they are in-existent is left quite unclear by Harman (as he admits, 666 ff.). I do not comment directly on this view in this dissertation, but must admit I find it quite implausible when construed of phenomenal properties (rather than, say, perceptual beliefs).)

What, for example, of someone who says merely that, for A to be in state which we would describe as “apprehending the quale q” is for the brain of A to be engaged in a certain kind of neuronal (or higher-level) activation? From this it certainly does not follow that q denotes a property of A’s brain … but that is because this claim alone simply does not specify what sort of property q is; it does not include a complete account of the denotation of q. If we press the matter, the live options seem still restricted to one of our four. For example, one might hold that talk of “apprehending the quale q” is just a figure of speech, and that really there is no such property as “the quale q.” Or one might assert that the quale q is a property of perceived objects in the external world—that A is apprehending the actual whiteness of the milk, for example—or conceivably of phenomenal individuals intermediate between A and the external world, such as white sense-data. Finally, one might adopt the theory that the ‘q-ness’ of A’s apprehension has to do with some property of her brain—that, for example, to be an instantiation of quale q (or to be an apprehension which it is “like q to have,” or something of the sort) is just to be an instantiation of some functional property of brain-like systems.

Karl Popper and John Eccles’ The Self and Its Brain (published in 1977) also famously endorsed dualism; Arthur O. Lovejoy’s The Revolt Against Dualism (1960) is a classic treatment.

Though below I present my model of qualia, initially, as a reformulation of sense-datum theory, I do not endorse sense-data as phenomenal individuals—instead, I recast them as phenomenal properties.
free from most eliminativist objections; and I consider Dennett’s main arguments in more detail in section one of Chapter Eleven.

The endorsed option b), however, contains a great number of widely different positions. We can begin teasing these apart with the following question: Are qualia-types reducible to, or identical with, properties (of states of brain-like systems) which are apprehensible from the third-person—from an ‘objective’ point of view?¹⁰ For example, are qualia (such as, for example, the property *phenomenal redness*) reducible to, or identical with, certain functional or neuronal properties? On this point the more intransigent core of Hard Problem exponents—I shall call them “hard liners,” for short—differentiate themselves from other theorists, whom we might call the “reductive physicalists”:¹¹ the hard liners, of course, deny the possibility of such reduction; the reductive physicalists embrace it.

Let’s consider the reductive physicalists first, still an extremely disparate group but one we can organise into three categories according to their answers to the following third diagnostic question: With what general kind of objective¹² property are qualia types to be identified? There are three typical answers:

- c) Information-processing properties (such as indicating, within some more or less specific type of information-processing system, that an external object is red);
- f) Complexes of already known causal-physical properties (such as being a brain state with certain chemical and electrical attributes);
- g) As yet undiscovered causal-physical properties (such as, perhaps, being a certain sort of ‘quantum gravitational collapse’ across a set of neurones).

---

¹⁰ This kind of question is often phrased as simply “Are qualia reducible to (or identical with) properties of states of brain-like systems?” However, this wording is unhelpful: given option b), qualia are properties of states of brain-like systems, and so, of course, must be both identical with and (in some sense) reducible to themselves. What we want to get at is the kind of properties of brain-like systems that they are.

¹¹ Often, members of both parties consider themselves “naturalists”—Chalmers, McGinn and Searle are enthusiastic naturalisers of consciousness, for example, but none of them are reductive physicalists.

¹² Read, in this context, third-person accessible properties, apprehended through something like the (present or future) methods of physics.
The multitude who endorse e) can be further sub-divided into, for example, black-box behaviourists, classical (or Turing Machine) functionalists, representationalist (or teleological) functionalists, Higher-Order Thought theorists, connectionists, and neurocomputationalists. Takers for f) are less thick on the ground these days. David Armstrong's *Materialist Theory of the Mind* (1968a) is usually cited as the best extended treatment of this form of identity-theory, yet as we shall see it is not at all clear that he now holds view f) about qualia (Armstrong and Malcolm 1984), if indeed he ever did. Christopher S. Hill's book *Sensations* (1991) is, as far as I know, the only recent book-length defence of any prominence of a view like f), though there are probably some broadly reductive physicalist thinkers who hold that

---

13 If there are any left today; Ryle 1949 is a central example from the recent history of philosophy. Black-box behaviourism was the view that mental states can be identified with dispositions to behave in certain ways—for example, the sensation of redness consists in (among other things) the disposition to assent to the claim that a ripe strawberry is red.

14 Putnam 1967 is the classic source, but he no longer holds this view; Lewis 1966 and Block and Fodor 1972 were also influential. See Sober 1985 for more on the Turing Machine/teleological distinction. Classical functionalists hold that mental states are defined by their functional role as the effect of certain inputs and the cause of changes to other mental states and behaviour.

15 Dretske 1995, Lycan 1996, Tye 1996. Representationalist functionalists typically argue that conscious states are identical with representational states—thus, the sensation of redness is identical with any state indicating (in the right way) that something is red.

16 Rosenthal 1986, 1996. HOT approaches hold that phenomenal consciousness consists in having higher-order mental states directed at one's own first-order mental states; one must not only code the information that something is red, but also represent that information to oneself.

17 McClelland et al. 1986, Smolensky 1987. Connectionists treat the mind as a processing network, and mental states as states of that network; the sensation of redness, then, might be (an aspect of) the highly distributed state the brain “relaxes” into after being stimulated by a perception of something red.

18 Churchland and Sejnowski 1992, P.M. Churchland 1995, Clark 1993. Neuro-computationalists hold that brains are essentially computational systems, whose physical properties are highly relevant but only insofar as they serve the processing of information (including efferent signals to the body); the sensation of redness, for example, might be a “vector coding” in the brain. A progression of, arguably, increasingly sophisticated versions of the same instinct can be traced from behaviourism, through functionalism (which went “inside the black box”), to the “computer models of the mind” so popular in the 1970s, and now to an increasing focus upon the results of neuroscience.
Chapter 1: “Introduction.”

qualia are not identical with informational/functional properties, and who may therefore think
of them—by a kind of default—as identical with physico-chemical properties.\footnote{Chalmers writes of \textit{Don't-have-a-clue} materialism: “I don’t have a clue about consciousness. It seems utterly mysterious to me. But it must be physical, as materialism must be true” (1996a, 162). He suggests that such a view is widely held, but rarely acknowledged in print. This is one of eight possible views on phenomenal consciousness, according to Chalmers (1996a, 161–168); he rejects them all except for this one and his own (Naturalistic Dualism), and he prefers his own theory over \textit{Don't-have-a-clue} Materialism on the basis that it is by far the more worked out. My dissertation can be construed as an attempt to move forward on the cashing out of \textit{Don't-have-a-clue} Materialism without collapsing into one of the other categories, such as reductive functionalism.}

As far as option g)—as yet undiscovered causal-physical properties—goes the field is just as thin: the names of Roger Penrose and Stuart Hameroff are the main ones associated with it (Hameroff 1994, Penrose 1994, Hameroff and Penrose 1996).

The hard liners, by contrast with all three types of reductive physicalists, argue that we have \textit{a priori} reason to believe that qualia-types can never be reduced to \textit{any} of these objective properties. This arena is where most of the philosophical battles over qualia have taken place in recent years. Virtually all of the central thought-experiments and arguments in recent qualia-theory—such as inverted qualia,\footnote{All functional (or physical) properties could remain the same while qualia spectra (such as phenomenal colours) are inverted (Lycan 1973, Shoemaker 1982, Block 1990, Horgan 1984c, Chalmers 1995b, and so on).} the possibility of zombies (absent qualia),\footnote{All functional (or physical) properties could remain the same while qualia are completely absent (Kirk 1974, Shoemaker 1975, Block and Fodor 1972, Block 1978 and 1980b, Chalmers 1995b and 1996a, etc.).} the chauvinism/liberalism set of problems for functionalism,\footnote{Into this class I would put several well-known thought-experiments, such as Searle’s Chinese Room (1980) and Lewis’ mad pain and Martian pain (1980). The basic idea is that functional specifications either attribute mentality (or, for our purposes, qualia) to systems which intuitively cannot possess it (e.g. the economy of Bolivia), or deny mentality (qualia) to systems which do (e.g. radically insane humans).} the knowledge argument\footnote{One could have complete knowledge of all the objective facts and still lack important knowledge of qualia—what it is like to see yellow, for example (Jackson 1982 and 1986, Nagel 1974).} and the explanatory gap argument\footnote{No imaginable physical or functional explanation could help us understand why a certain state subjectively feels a certain way (Levine 1983; this argument is sometimes described as an epistemological version of Kripke’s (1971, 1980) modal argument against the identity theory).}—are designed either to show that any hypothesised correlation between
informational or physical properties and qualia may in principle break down, and thus that no identification is possible; or to show that any functional or physical theory must be inadequate to explain qualia.

Assuming with the hard liners for the moment that qualia-types cannot be reduced to or identified with objective properties, we can now ask them our final diagnostic question: Are these irreducible qualia mental or, in some sense, physical? I suggest below that it is not quite clear what the force of this question really is (what is meant by “mental” or “physical” here?) but the question does in practice allow us to discriminate four varieties of hard liners:

h) Qualia are ‘physical’ because qualia tokens are (always) physical (Token-identity physicalists);

i) Qualia types are ‘physical’ because they supervene in the appropriate way upon the physical (Supervenience physicalists);

j) Qualia are distinctively ‘mental’ properties (Property dualists);

k) Qualia are neither ‘mental’ nor ‘physical’ (Neutral monists or panpsychists).

View h) consists in holding that qualia types are neither reducible to nor identifiable with objective types, but that each and every token quale is identical with some physical token—that is, since we are talking about properties here, some physical property-token. As far as I can tell, this is not a very widespread hard line view when it comes to qualia (though it is one that has a fair number of takers with respect to cognitive mental states such as the propositional attitudes). It is possible that John Searle (1992) holds a view somewhat like this, since he denies that he is a property dualist but is also sceptical about the physicalist supervenience of qualia; Sydney Shoemaker also sometimes seems to be making claims of sort h) (e.g. 1982). Supervenience physicalism, i), is much more widely discussed: central sources on this notion include Hellman and Thompson 1975, Fodor 1974, Horgan 1982, Post 1987 and Marras 1993. Joseph Levine (1983, 1993), Brian Loar (1990) and perhaps Owen Flanagan (1992) apply supervenience

---

25 Davidson's anomalous monism (1980) is the central account (but with respect to events, not properties) ... though it is also possible to construe his theory as a form of neutral monism.
physicalism specifically to qualia. Option j), property dualism, is usually based upon the claim that qualia are such a radically different kind of property than all currently accepted physical properties that they cannot count as physical—typically, this is because it is thought that qualia could not possibly be described within the language of current or future physics. Contemporary adherents include Campbell (1970), Honderich (1981), Jackson (1982), Nagel (1993), Sprigge (1994) and perhaps Chalmers (1996a). Exponents of view k) usually hold that what we call ‘mind’ and ‘matter’ are two aspects of a more basic substance or reality. David Chalmers is tempted by this view (1996a), and Michael Lockwood’s *Mind, Brain and the Quantum* (1989) defends the notion that experience is the intrinsic nature of the physical.26

Now, where does this dissertation fit into the sub-structure I just outlined? Which of positions e) to k) does it elaborate and defend? The simplest answer to the latter question, for good or ill, is “none of them.” The position taken is rather the following: that there are very strong grounds to believe, even at this early stage of the research game, that qualia are phenomenal properties of the brain (in a quite full-blooded sense of “phenomenal”); but that the case is not even close to being closed on whether these properties are identifiable from the third-person standpoint—or, to put it another way, whether qualia are identical with “objective” properties of the brain. This, in my view, is at base an empirical matter, and not purely a philosophical one: it is a question that is best answered, not from the armchair, but in a laboratory examining the brain and its properties.

On the question which distinguishes the hard liners from the reductivists, then—“Are qualia reducible to objective properties?”—I have to answer that I do not know … nor do I think anyone really knows. More positively, I do not think that the a priori arguments from the hard liners carry the day (and I provide some arguments for this claim in Chapters Eight and Ten, as well as some considerations in favour of metaphysical—as opposed to merely epistemological/pragmatic—reducibility). Further, one of the aims of this dissertation is to suggest that by far the most fruitful research paradigm for the empirical neurosciences, at this stage, will be to

---

26 Bertrand Russell (1927), Herbert Feigl (1958) and Simon Blackburn (1992) have also countenanced this idea.
assume qualia are reducible to empirically accessible properties and to try and find out what these might be.

On the other hand, this dissertation takes the stance that it is not legitimate to pursue this project of identifying qualia with objective properties by deflating qualia—by ignoring the first-person phenomenal aspect of qualia. Any plausible empirical research project must hold constantly in view the fact that qualia are the colours, the smells, the tastes, the pains, the emotions of our mental life. When looking at a glass of milk, the hard liners are right to insist that it is the tokening of the property of phenomenal whiteness that is the explanandum, and not merely the information completely expressible by the proposition “the milk is white.” Thus, although qualia may be identifiable with objective properties, they will appear different from the first-person and third-person perspectives. What is needed, ultimately, is a theory which will explain how an objectively describable property of such-and-such a type feels a certain way to the subject within whose brain it is instanced.

This dissertation’s sympathies lie most strongly, therefore, with physicalist option f): that qualia are identical with physico-chemical properties of the brain. The assumption that reductive physicalism is possible conflicts with all the hard line views, and the emphasis upon the phenomenal nature of qualia tends to come into conflict with the information-processing approaches (e). I do not argue in this dissertation that functionalism or computationalism with respect to qualia are wrong in all their varieties, but in section two of Chapter Eleven I do make phenomenality-based criticisms of Dretske’s representationalist functionalism which are probably applicable to other versions of functionalism.27 This dissertation remains neutral on

---

27 The making of qualia-related objections to functionalism has become virtually a philosophical cottage industry in the past twenty-five years or so, and the ground is now so well-trodden that I feel I have little fresh to say on the subject. I think it is more important, at this stage, to flesh out a fairly detailed and plausible phenomenalist account of qualia to provide a more solid foundation for those objections (which so far have not been conspicuous for their success).

It is important to note that, though this dissertation argues that qualia are phenomenal properties—they are felt whiteness, bitterness, pain, etc.—extra arguments would be needed, which I do not provide, to focus this account as an attack on functionalism. In particular, I do not argue explicitly that phenomenal properties are ‘intrinsic’
positions of type g), that some quite radically new physical theory (such as a theory capable of predicting and explaining the ‘collapse’ of the quantum-mechanical wave function) will have to be developed before qualia can be reductively explained.\(^{28}\) I feel the pressure of the intuition that qualia cannot be explained in terms of sodium ions and dopamine; however, in my view, not only is the jury not yet decided on the issue of the new sciences, the trial is not even ready to begin.

Though I am not a hard liner (in virtue of holding open the possibility of qualia reduction), this dissertation does accept that one of the four hard line positions may be true—that is, reductionism may be true, but as far as we know now it may also be false. Of the four hard line options, I find supervenience physicalism the most plausible; this option (i) is discussed at some length in Chapter Eight, where I argue that it is possible to hold a minimally physicalist position based merely on the claim that qualia supervene appropriately on the (objectively) physical. I do not deal with token physicalism (h) explicitly in this dissertation, but much of (whatever that is supposed to mean), or, more concretely, that they are non-relational properties, best described by monadic predicates. Thus it probably remains in principle open to the functionalist to agree with the conclusions given here, but assert that some kind of relational account can be provided for phenomenal whiteness. Care should be taken, however, to ensure that this functionalist response does not just change the subject: it would not be enough to simply proffer a rival theory in which what relevantly happens in the brain for a normal subject seeing white is entirely relational in character; instead, to meet rather than ignore the conclusions of the dissertation, the functionalist would have to show that phenomenal whiteness—that property which is tokened during certain cases of conscious visual sensation … the occurrent colour which appears to cover the surface of white objects—is relational. Though it is very hard to express the difficulty clearly (more efforts will be made in Chapter Three), this would appear to be a much more difficult task for the functionalist.

\(^{28}\) The major players in this market—Penrose and Hameroff—base their rejection of current physics upon interpreting Gödel’s incompleteness result as showing that human conscious thought is sometimes sound yet non-algorithmic, while they claim that current science recognises only algorithmic processes (or absolute randomness, which is no help). They have found relatively few supporters and been quite bitterly attacked in the literature (e.g. Putnam 1994, Grush and Churchland 1995); discussion of their ideas at 1996’s Towards A Science of Consciousness II conference at Tucson was, perhaps surprisingly, highly negative. The criticisms, however, focus upon Penrose’s interpretation of Gödel, and the empirical support for the usefulness of Hameroff’s quantum microtubular theory in solving problems to do with consciousness (though I would have thought it was far too early in the day to be very demanding about the latter).
what I have to say in Chapter Ten regarding the metaphysical peculiarity of irreducible patterns is relevant—that is, I emphasise quite how radical it is to assert that an upper level type corresponds to no lower level type at all (even an ‘unnatural’ one), though of course I do not prove \textit{a priori} that such a thing is empirically impossible. Chapter Ten is also pertinent to the question of property dualism (j), insofar as I suggest that whether we were to call irreducible qualia “mental” or “physical” is far less important than the issue of whether they \textit{are} in fact irreducible or not; Chapter Eight is relevant since I try to show there that qualia could be irreducible and “physicalism” still be true; and Chapters Four to Six are significant insofar as they lessen the air of special epistemological mystery surrounding qualia. I do not consider option k), neutral monism, in this dissertation, though it is probable that any considerations I put forward which lessen the pull of property dualism will also weaken the claims of neutral monism.

Before moving onto more substantive matters, perhaps now is a good moment to briefly sketch the structure of this dissertation.

2. OUTLINE OF THE DISSERTATION

Chapters Two and Three try to improve upon long-standing arguments and intuitions to provide what is tantamount to a proof that qualia exist and are non-identical with properties of external objects. This is a general refutation of eliminativism and externalism, and bedrock for the qualia theorist. I argue that qualia are in fact properties of states of the central nervous system (CNS). I make this claim on the basis of a reformulation of the argument from illusion that relies upon the premise that there are no unowned properties. This version of the argument establishes that, for example, when I have the sensation of an orange light floating in front of a blue wall (an after-image), then the property of \textit{orangeness} that is being tokened—since there \textit{is} phenomenal orange “going on”—is not a property of any external object, and so (via a brief rejection of other candidates like phenomenal individuals) is most plausibly a property of my CNS. It follows from this, I try to show, that the instances of, for example, the quale orange are not things that are orange, but are phenomenally orange sensations: that is,
parts of the brain are (obviously!) not actually coloured orange, but parts of the brain do have the property of being phenomenally orange. Some brain states are sensations of orange (that is, of the ‘real’ colour orange, a property at least putatively possessed by external objects), and among the properties of (at least some of) these brain states is that of phenomenal orange-ness—of, roughly, ‘feeling orange.’

I then proceed to discuss the epistemological and metaphysical consequences of this view of qualia: being forced to claim that occurrent sensory properties are not external (and are not to be modelled on external properties, such as ‘actual’ colours and shapes) but are instead phenomenal is, it turns out, not so much a problem as an opportunity. In Chapters Four through Six I examine some of what have been thought to be the major sticking points in theorising about qualia, under the general rubric of the epistemology of qualia. These include notions like the special ‘privacy,’ ‘subjectivity,’ ‘incorrigibility,’ ‘direct apprehension,’ and so on, of qualia. I analyse and clarify the important notions involved with immediacy, certainty and privacy (such as indubitability, incorrigibility and privileged access), and describe the manner in which the apprehension of qualia has these characteristics. I conclude that experience does indeed form a rather different epistemological ball-park than the rest of the world; however, I strive to show that these epistemological characteristics are not nearly so bizarre as to motivate eliminativism (contra Dennett) or dualism (contra Nagel or Jackson), and try to demonstrate that theorising in this area is perfectly coherent and possible.

In Chapters Seven through Ten I consider the metaphysical relation between qualia and the brain. I emphasise that discovery of the nature of the dependence of qualia upon the brain states of which they are properties is ultimately an empirical matter; however I develop a schema for the supervenience relation and use this to define just what it would be for the thesis of nonreductive physicalism to be true with respect to qualia. I then argue that considerations that purport to show a priori that qualia are physical or not (such as appeal to the logical possibility of inverted or absent qualia, or alternatively the claim that epiphenomenalist accounts of qualia must be bankrupt) are not successful. Finally, however (Chapters Nine and Ten), I consider what it would be to claim that qualia are “at a higher level” than the brain but
irreducible to base-level properties of the brain, and suggest that this is a more metaphysically uncomfortable position than is always realised.

Lastly, Chapter Eleven contains discussion of the major arguments of three central proponents for three accounts of qualia which are importantly different than mine: Dennett on eliminativism, Dretske on externalism, and Chalmers on property-dualism.

For concreteness and convenience this dissertation focuses upon sensory qualia (like colours and smells), and tends to ignore other varieties of qualia, such as those that might go with emotional states, or perhaps the calling to mind of beliefs. I have also focused selectively within the literature, dealing mostly with that on the qualia-analogues of the philosophy of this century rather than casting further afield in the history of philosophy or doing more than glance at work in other disciplines, such as the debate on mental imagery in empirical psychology. I have tried to remain sensitive to the most significant recent results in the empirical sciences of the mind-brain, but have not considered this literature exhaustively, and generally do not rely upon empirical details to support my conclusions.

3. TERMINOLOGY MATTERS

Since in the next two chapters I will be refurbishing arguments used during the sense-datum debate of the 1920s to the 1960s, it is necessary to consider the relation between the terminology of that time and the “qualia” terminology used today. This is especially important since, in my view, the discussion of sense-data was significantly muddied at the time by inadequate attention to the distinction between phenomenal properties and phenomenal individuals, and because the modern discussion of qualia is, I believe, bedevilled by unnecessarily strong assumptions about the term. My position is this: both the terms “sense-datum” and (as it is often used today) “quale” are multiply ambiguous, in roughly the same way. I shall therefore supersede both ways of talking with my own, hopefully less ambiguous, terminology (and will recast the

---

29 Owen Flanagan (among others) holds that “all and only conscious states are qualitative [by which he means, roughly, qualia-laden]” (1992, 72), and (for at least one important usage of “conscious”) I tend to agree.
sense-datum arguments in the new language).

First, I take it that the general domain within which both the terms “sense-datum” and “ quale” are intended to operate is that of the *phenomenal*, or felt, or qualitative, or *how it is like something to be*. Provisionally, a useful way of understanding this talk is by allusion to a roughly Kantian (or classical Greek) notion of phenomena: objects, properties or events *as they appear in experience* are what we label “phenomenal;” as they are ‘in themselves’ they are “non-phenomenal.”

Thus, at least loosely speaking, the greenness ‘of’ some visual sensation is phenomenal (while the ‘actual’ colour of an unperceived spruce tree is not). It is important to note, however, that this terminology is not by itself enough to commit us to a distinction between *types of properties or objects*. The phenomenal/non-phenomenal distinction, as it is used here (unlike, I believe, in its properly Kantian analogue), is supposed to be neutral between classifying *types of properties or objects* and distinguishing between something like ‘*modes of presentation*’ of properties or objects. For example, as far as the terminology goes, phenomenal greenness—greenness as it is experienced—might be a different property than non-phenomenal or ‘actual’ greenness, or it might just be the very same property when ‘presented to experience.’ (In Chapter Three, we shall move beyond this insipid neutrality, but thus far I am simply attempting to establish minimal definitions of my terms.)

Within the class of the phenomenal, the most important distinction to build explicitly into our terminology is that between phenomenal *properties* and phenomenal *objects*. I shall henceforth restrict the term “qualia” (unless otherwise noted, and when not qualified by scare-quotes) to *phenomenal properties*—the taste of peanut butter, the softness of a feather’s touch, the greenness of a visual sensation, and so on. (I take it that this is by far the most common usage for the term in the contemporary literature, and so restricting the term only to properties is not

---

30 This usage also respects the emphasis upon appearances, contrasted with reality, in *phenomenalism* and *phenomenology.*

31 In addition, I intend to presuppose no particular theory of perception—such as an indirect theory—by the use of this terminology. So phenomenal greenness, as far as the words go, could merely be actual greenness directly perceived.
unduly distorting.\textsuperscript{32} At this point, once again, the relation—identity or non-identity—between phenomenal properties and the objective properties of external objects is left unspecified. For example, again, the greenness we experience when we look at a summer lawn is a quale; it may, or may not, be identical with the objective property greenness possessed by the lawn.\textsuperscript{33} So I do not consider it part of the meaning of the word “qualia,” in my usage, that they are mental properties, or that they are intrinsic properties of conscious states, or that they are atomic, or that they are not reducible to “physical” properties … things like these may or may not be true of qualia, but that remains to be discovered; they are not entailed by (my) mere use of the term. All I mean to pick out with the word are colours as they are presented to us in visual sensation, tastes as we experience them, sounds as we hear them, and so on.

To distinguish talk of qualia from talk of objective properties, I shall prefix qualia names with the word “phenomenal.” Thus “phenomenal redness” is the redness involved with our perception of a ripe strawberry; “redness” (or, for extra clarity, sometimes “external redness”) is a property possessed by that strawberry.\textsuperscript{34}

Qualia are plausibly to be distinguished from experiences (here construed as products rather

\textsuperscript{32} Thus, for example, the Cambridge Dictionary of Philosophy defines “qualia” as “those properties of mental states or events, in particular of sensations and perceptual states, which determine ‘what it is like’ to have them” (Audi 1995, 666, my italics) and Ned Block in the Blackwell Companion to the Philosophy of Mind describes them as “phenomenal properties of sensations, feelings, perceptions and, in my view, thoughts and desires as well.” (Block 1994, 514, my italics). Even Daniel Dennett, in his influential characterisation of and attack upon qualia in “Quining Qualia,” assumes qualia to be a (to him mythical) species of “properties of conscious experience” (1988, 42). Examples such as these can be multiplied almost indefinitely. This view is not quite universally held, however; for example Own Flanagan defines a quale at one point as “a mental event or state that has, among its properties, the property that there is something it is like to be in it” (1992, 64). In my terminology, described below, these events or states would be called either “experiences” or “qualifiers,” depending on what exactly is meant—visual images of speckled hens are experiences; events or states that have qualia among their properties are qualifiers.

\textsuperscript{33} A. R. Lacey’s dictionary definition of “quale,” likewise, explicitly leaves open the possibility of this identity: “A quale, such as red, is a quality considered as it appears to consciousness rather than as science might define it” (1996, 278).

\textsuperscript{34} Obviously, given the previous paragraph, this language is not supposed to carry any presupposition to the effect that phenomenal redness is a different property than external redness … though I shall argue presently that this is in fact the case.
than processes)—from what might be called sensations or images or regions of the sensory field. Thus, for example, visual images of spreading elms, the experience of the first movement of a Beethoven symphony, or the perception (using sight, smell, hearing and touch) of a wet dog, are probably not “qualia” in my sense since such things are not typically treated as properties but as a sort of individual. I shall use “experience” as a term of art for such phenomenal individuals, including famous entities like Moore’s white envelope or Ayer’s imagined speckled hen. However, to avoid prejudging any ontological issues, I shall take the word “experience” to carry as little metaphysical baggage as possible: thus, I shall treat experiences as being nothing more than structured sets of qualia-tokens. For example, the visual “experience” of a white envelope consists, as far as we are concerned, merely in the bundling, in a certain way, of tokens of the quale phenomenal whiteness with that of phenomenal rectangularity, and so on. The experience of a wet dog combines qualia for shape, colour, movement, smell, touch, sound, etc.. That is, when someone experiences a wet dog, what that person experiences is shape qualia, colour qualia, smell qualia, etc., concatenated in a certain way: once all these experiential properties have been enumerated, there may be nothing else—no additional wet dog image, and certainly not necessarily one in which the qualia inhere. In other words, experiences are not to be

---

35 When used as a count noun: “an experience of a wet dog” involves this ‘technical’ usage, whereas talk of “one’s visual experience (generally)” does not.

36 To quote C.I. Lewis (in a similar but not identical context): “It seems better to use language which is familiar, even at the risk of some ambiguity, than to invent a technical jargon which would, after all, be no less ambiguous until its precise connotation could be discovered from its use” (1929, 57).

37 The model here is similar to that of the “bundle theory” of concrete external objects, as espoused by Berkeley (1710), Russell (1940) and Goodman (1966), whereby concrete objects are exactly sequences of complexes of co-instantiated properties; it differs, however, in remaining neutral upon whether the bundles of properties picked out by the term “experience” might also have some “core” of “substance” with, perhaps, other properties as well.

Really, all I want is to continue to be able to talk about speckled hens, qua images and not actual domestic fowl out in the world, without being committed to ghostly chicken-like individuals, pictures in the head, etc. But I want to do more than just talk about “apparent chicken-sensings,” or believing that there is a chicken out there of a certain sort, or being disposed to do such and such—I want to talk about the image itself. The term “experience” is my solution to this problem, not a fully fledged (so to speak) theory of mental imagery. It is supposed to avoid presuppositions, rather than introduce new ones.
taken as ghostly individuals—little mental analogues of envelopes and dogs; rather, they are simply “chunks” of the stream of consciousness.\footnote{That is, perhaps, the ordered sets of qualia which are experiences are themselves members of a temporally ordered set of sets (for some individual).}

It is important to bear in mind that qualia are not properties of experiences, on this usage; instead, experiences are, so to speak, built up out of qualia.\footnote{Experiences can even be treated as complex qualia-tokens—instantiations of the complex phenomenal property of being (the experience of) a wet dog, say. I avoid this way of talking because it seems to me to blur the perfectly good intuitive distinction between phenomenal properties and phenomenal objects—between a token of phenomenal greenness and the experience of a summer lawn, for example. As we shall see, the sense-datum theorists often argued for non-identity between external objects and mental individuals, rather than between objective and phenomenal properties—we shall therefore find some kind of bland characterisation of a mental object to be of great use in reconstructing their arguments. Further, I am prepared to be highly nominalistic towards experiences—some particular stretch of a stream of consciousness might, it seems to me, be divisible into experiences in a number of different, equally sound, ways, for example—but might in the end want to be much more of a realist about qualia. If the reader prefers, however, it does seem possible to collapse talk of qualia and experiences into talk of more and less complicated phenomenal property-instantiations without harm to the substance of this and later chapters.}

So, for instance, we will insist that, strictly speaking, the experience of a speckled hen does not have the property of being phenomenally brown—rather, it is partly constituted by the tokening of phenomenal brownness.\footnote{It might be that one would want to talk about “phenomenal experiences” in more full-blooded terms than I do here: as the entities of which qualia are properties, for example, which perhaps have causal powers, location, the ability to subsist independently of the perceiver, or whatever. My terminology does not rule out this way of thinking; it merely requires that one not use the word “experience” of such entities, since that term is restricted to mere property-bundles; instead, individuals of this kind are a species of qualifer (see below)—a sort of thing which has phenomenal properties.}

Of what, then, are qualia properties? As we have seen, this is not a terminological question but one of substance, to which the next two chapters are mainly directed. For now I will simply introduce a neutral term—“qualifiers”—for whatever it is of which qualia are properties.\footnote{Thanks to D. F. Hamilton for suggesting this term.}
We may now state straightforwardly that by “experience” I mean something which is not a quali-fer, and that my claim in Chapter Three will be that external objects are not quali-fers. There has historically been a notion of phenomenal objects as expressly being something like non-physical entities that exist intermediately between “us” and the world: I take this possibility to be peripheral to the contemporary debate, but will discuss it briefly in Chapter Three. I shall call this the doctrine of the ghost quali-fer.

Perhaps a summary of the (terminological) distinctions I shall be relying upon is of use:

1) The furniture of the external world—its properties, objects, forces, and so on. I shall typically demarcate this sort of entity with the categorising predicates “external,” “objective,” “concrete” and the like. For example, “external greenness” is simply supposed to denote some property held in common by unripe oranges, pine leaves and some book covers in virtue of which we call them all green.

2) Qualia: phenomenal properties (which, as far as the word itself as it is used here goes,

---

42 Or at least, that they do not instance the phenomenal properties we undergo during perception, though I suppose they may still be qualifers

43 Note, however, that the terminology I shall use is consistent with this thesis, since one possible theory in this area is that qualifers are indeed some sort of non-physical perceptual intermediaries. So, if we suppose (as some do, though I am not entirely willing to go along with them) that, for example, the term “sense-datum” in the first two thirds of this century was always expressly meant to pick out non-physical, intermediate entities, then my interest in the sense-datum literature will only be whether it helps to establish the weaker claim about qualifers whatever they are, without any theoretical preconceptions.

44 A listing of kinds of ontological status can also be read off from this list. Thus, I am minimally ontologically committed to the furniture of the external world (whatever it is), including properties, objects and forces, and also a similar set of “internal” furniture (see below for an account of the word “internal”—all I mean by it here is, roughly, not being spatially located outside the central nervous system): internal objects (such as neurones and their groupings); internal properties (that is, properties of the above objects); and internal forces (fields of force which can be said to be internally located—i.e. perhaps those whose origin is internal, though not all the field need lie within the CNS). I will be arguing in the next two chapters that qualia are internal properties (and, hence, that qualifers are internal objects).

45 The phrase itself is supposed to carry no presuppositions about whether there is indeed such a property; whether, if it exists, it is to be identified with phenomenal greenness or not; whether the property is intrinsic, relational, or what; and whether or not it is the property we intend to pick out with the word “green.”
may or may not be identical with external properties\(^{46}\) … though I will argue they in fact are not so identical).

3) Experiences: sets of qualia bundled together in a structured way to make phenomenal individuals such as visual images like those of white envelopes or speckled hens, or sensations unified in some way through time (such as the experience of a symphony). Experiences are not qualifiers.

4) Qualifiers: individuals which possess—in which inhere—phenomenal properties; that is, qualia-holders.\(^{47}\)

Before moving on, just a couple more definitional points. I adopt the following way of talking, which I take to be fairly standard: property types are “instantiated” (or “tokened”) as property tokens, and these property tokens “inhere in” (or are “instanced by”) substance tokens—or, more neutrally, “individuals”\(^{48}\) of some sort. For example, the property of whiteness is instantiated in a property token of whiteness which inheres in my coffee mug. In the noun form, I shall treat “property tokens” and “property instantiations” as being synonymous—both meaning token occurrences of some universal—and will use them interchangeably for variety’s sake. By contrast, by an “instance” of a property I shall always mean the individual which supports some property token. Thus, my coffee mug is an instance of whiteness, and that

---

\(^{46}\) The term “quale” was, I believe, first introduced in the nineteenth century by the American pragmatists (such as Peirce and James) who meant it to be neutral with respect to being physical or mental. However, C.I. Lewis, who helped popularise the term in this century, explicitly took the view from the outset that “although such qualia are universals, in the sense of being recognized from one to another experience, they must be distinguished from the properties of objects. Confusion of these two is characteristic of many historical conceptions, as well as of current essence-theories” (1929, 121).

\(^{47}\) One might be wondering at this point where things like mental content, the propositional attitudes, or the unconscious fit in with this taxonomy. The answer is that it is neutral with respect to them. For example, as far as I know nothing whatever is implied by this taxonomy, or the conclusions that will be built using its terms, about the nature of representation—the way in which certain mental states (plausibly, in at least some cases, including qualifiers) come to have certain contents. Thus, for example, one might want to say that the visual image of a tree represents that tree because it is caused by it in a certain way: fine.

\(^{48}\) I take this to be roughly equivalent to—though perhaps even more vague than—the term “objects,” as I have used it above.
whiteness is an instantiation of the property.

Next, a word on “internal” and “external.” The distinction is usually used in this context, I take it, to demarcate the contrast between the “mind” (and what “takes place in the mind”) of the perceiver, and the so-called material world. However, this usage illicitly begs too many questions about, for example, the mental-physical distinction, for my purposes, and seems unnecessarily obfuscating. By “external” I shall here simply mean that part of the physical world external to the central nervous system of the experiencer⁴⁹: so, a house is “external,” area V1 of the visual cortex is “internal”; my left foot is “external,” my spinal cord is “internal”; and so on.⁵⁰ Note that Cartesian souls or whatnot, should one insist on hypothesising them, still on this usage count as “internal” (even if they were somehow spatially located “outside” the CNS), since they are (“mental” and thus) not part of the external physical world.

Third, in what follows I will sometimes have occasion to distinguish between the processes of perception and sensation. “Perceiving” is a success term for a process which requires the more or less accurate picking out of some object in the external world; “sensing,” by contrast, is a broader term which I shall take to refer to all cases that might (though they need not) be mistaken for perception, and which thus does not presuppose the existence of any actual external object resembling that sensed.⁵¹ We can use the terms “perception” and “sensation” to

⁴⁹ One complication, however, is that I do not wish to assume all possible experiencers must have central nervous systems: I should therefore be understood as talking about CNSs or their ‘equivalent’—which, I confess, does not seem at all easy to cash out at the present time—or as tacitly restricting my claims to the class of, say, higher vertebrates, in the hope that they can later be generalised further. My primary interest, in any case, is human phenomenal experience.

⁵⁰ The Oxford Companion to the Mind defines the central nervous system as “the physical substance that provides its possessor with genetically determined ways of behaving and also ways of changing this behaviour” (Nathan 1987, 514). In vertebrates it includes both the grey matter and the white matter of the brain and the spinal cord. Including as this does such areas as the medulla oblongata (the slightly enlarged area at the top of the spinal cord, at the base of the brain) or the corpus callosum (the bundle of fibres connecting the two cerebral hemispheres) it might be thought that one could restrict the location of qualia even further than merely to states of the CNS—however, given the diffuse nature of mental states, I think it better to risk including too much of the world than too little.

⁵¹ This distinction is a fairly common one (or at least it used to be); it can be found, for example, in Jackson 1977,
talk about the ‘products’ of these processes—i.e. the experience (in the sense of the term defined above) so produced. That is, perceptions must more or less accurately ‘represent’ some particular external object; there is no such requirement upon sensations. For example, my visual experience of my computer monitor is (I hope!) the product of some process which is both a sensing and a perceiving; my visual experience of flashes of light when I rub my eyes is just a sensation and not a perception.\textsuperscript{52}

It is as well to note, incidentally, that I consider the “objects” of perception—the things perceived, or which we aim to perceive—as being always external objects, and never, say, experiences or perceptually intermediate ghost qualities. When I look about my room I see a computer, green plants, a bookshelf—it is, in my view, a malapropism (or at any rate a turn of phrase I wish to avoid) to say that I “perceive” experiences of my computer, or plant-related qualia (unless, that is, qualia are to be identified with the objective properties of objects, a position which as yet still remains open).\textsuperscript{53}

\textsuperscript{4, and Ayer 1940, chapter 1.}

\textsuperscript{52} Another way to draw the same distinction is to think of sensing as being the process of perception “from the inside,” so to speak—as opposed to the “objective” causal process of receiving information from the external world, which is perception.

\textsuperscript{53} It is sometimes difficult to avoid phrases—like “object of perception”—that have the ring of earlier accounts in which experiences are “objects” that we “directly” perceive, and from which we “infer” the existence of secondary, external objects of perception. I mean no such thing, and must ask for some charity from the reader in
Chapter 1: “Introduction.”

Now, having set the scene, onto more substantive philosophical matters: over the next two chapters I shall argue that CNS states have phenomenal properties ... that the redness tokened when one sees a ripe strawberry—redness as it is presented to us in visual sensation—is a property not of strawberries but of brains.

avoiding such interpretations. The relation between perceptual qualia and perceived objects is, of course, an interesting one, but it is not something on which I have much to say here (though see my comments on “directness” in Chapter Four). Generally, I suppose that we perceive the world “through the medium of” experience—that, somehow, experiences “represent” the external world (or at least is made up of phenomenal properties of the states that do represent the world).
Chapter 2: Arguments for ‘Sense-Data’

I will argue in this chapter and the next towards the claim that (sensory) qualia exist and are distinct from properties of external material objects. That is, I will argue that the qualia associated with sensation in human beings (phenomenal greenness, loudness, smelliness, etc.) are at least sometimes not properties of external perceived objects (trees, trumpets, rotten eggs).¹ They must therefore, I will suggest in Chapter Three, be properties of some other class of individual instead—most plausibly, states of the human nervous system. To defend these claims I shall bring back from the grave a series of arguments once used by (among others) the sense-datum theorists of the first half of this century—philosophers like Moore, Russell, Ayer and Price. Most important among these is a new and improved version of the argument from illusion, rumours of whose death, I shall try to show in the next chapter, have been greatly exaggerated.²

1. THE SENSE-DATUM THEORY

I take G.E. Moore’s position in Some Main Problems of Philosophy to be a central, or at least seminal, expression of sense-datum theory. There he argued that at least two ‘things’ are involved in the act of perception. When looking at a white envelope, he urged, it is certainly and commonsensically true that one sees it, that object, which is at any one moment “at some one definite

¹ Obviously, this is not a new idea: it can be found—at least with respect to the ‘secondary’ qualities—throughout the history of philosophy, from the Greek atomists and Plato, to Galileo, Descartes and Boyle, to the British Empiricists and beyond. In the recent history of philosophy, G.E. Moore (in his lectures of 1910–11, later published in Moore 1953) was probably the first twentieth century philosopher to place stress on the distinction between what he called ‘sense-data’ (a term he probably introduced) and external perceived objects.

² Arguably, this is one of the two great, influential arguments in the history of the philosophy of mind. The other is Descartes’ argument, today often called the modal argument from conceivability, which purports to show that the mental cannot be identified with the physical since it is possible for one to exist without the other. In the recent literature, the argument from conceivability has attracted by far the more attention (e.g. Chalmers (1996a) versus P.S. Churchland (1996)); I believe, however, that the argument from illusion (perceptual relativity) is in fact the more powerful, and I try make a start at showing this here.
place in space” (1953, 30). However, for Moore, there is another ‘thing,’ non-identical with the first, which is also involved in the perceiving of the envelope:

... a patch of a particular whitish colour, having a certain size, and a certain shape, a shape with rather sharp angles or corners and bounded by fairly straight lines. (1953, 30)

He goes on to write that:

I propose to call these things, the colour and size and shape, sense-data, things given or presented to the senses—given, in this case, by my sense of sight. ... [S]o too when I hear a sound, I directly apprehend the sound; when I feel a toothache I directly apprehend the ache: and all these things—the whitish colour, the sound, and the ache are sense-data. (1953, 30–33)

By “sense-data,” at this point, it seems that Moore is thinking at least in part of what I am calling qualia—of properties like colour, size, shape and achingness. However, his intended contrast with the external perceptual object is most plausibly what I call experiences (i.e., roughly and at a metaphysical minimum, structured bundles of those qualia)—coloured patches, toothaches and sounds—and Moore also calls these “sense-data.” Probably the best reading of the thesis of the Moore of 1910, therefore, at least as described using the terminology defined in the previous chapter, is the following:

(1) Some external perceived objects are non-identical with experiences (in the ‘technical’ sense introduced above), and vice versa. For example, envelopes are not identical with visual sensations of envelopes.

That is, Moore attempts primarily to draw a distinction between two different types of individ-

---

3 Since, in Chapter One, I have proclaimed the term “sense-datum” ambiguous, it is now incumbent upon me to try and disambiguate the usage through the application of my own terminology.

4 Moore quite likely in fact meant much more by “sense-data” qua individuals than I mean by “experiences”—for example, he held that they are what is ‘directly’ perceived. All such matters I intend to bracket for the time being, however, and I offer this only as a very minimal construal of the real Moore.

5 By this, Moore does not simply mean that concrete individuals are not identical with bundles of properties; he must claim that, even if some external object were reducible to the set of its properties, that property-bundle would be non-identical with the relevant experience.
ual. However he could easily have chosen instead to have put forward the following claim about properties:

(2) Some qualia are non-identical with properties of external perceived objects, and *vice versa*. For example, the actual whiteness of an envelope is not identical with the quale of whiteness experienced by that envelope’s perceiver—actual whiteness is not the same property as phenomenal whiteness.

Or he could have made this claim:

(3) Some external perceived objects are non-identical with individuals in which qualia inhere (that is, they are not qualifiers), and *vice versa*. For example, envelopes are distinct from the individuals which have among their properties the qualia associated with perceptions of envelopes—envelopes are white, but they are not phenomenally white.

Some $X$ and $Y$ might be non-identical either because $X$ exists and $Y$ does not, or because both $X$ and $Y$ exist but $Y$ is not identical with $X$; non-identity claims like these three should be understood as making the latter assertion. Thus they include an implicit existential: experiences, qualia, or qualifiers actually do exist.

How are theses (1) to (3) related? Each of the three separately entails the truth of the other two, given the following two plausible assumptions: that property instantiations (i.e. property tokens—a particular token of redness, say) always inhere in—are the properties of—some in-

---

6 And not, as some familiar but loose characterisations might lead us to believe, between external individuals and sensed properties. Here, for example, is R.J. Hirst’s take on Moore: “We are urged to distinguish between (i) public physical objects, such as tables and chairs, cabbages and kings (or rather kings’ bodies), and (ii) the immediate data of experience, colours, shapes, sounds, smells and so on” (1959, 26). It is not tables and colours being contrasted, however, but tables and visual images of tables.

7 This need not—though it could—be the thesis that ‘objective’ whiteness and the quale whiteness are different *universals*; the minimal claim is that at least two property *instantiations* are involved.

8 To put it another way: when we see a white envelope (though, I shall want to say, the envelope is indeed the thing that is perceived and is actually white) the *phenomenal* whiteness we experience—the quale whiteness involved with our visual experience—is *not* among the properties of the envelope. Envelopes, that is, do not instance qualia.
individual; and that any quale instantiation can form part of an experience (e.g. that any token experienced shape or colour can be part of a visual image). Suppose thesis (1) above were true: then we have two non-identical individuals, \( a \) and \( b \), one of which is a bundle of qualia. Two different individuals cannot both possess the same property instantiation—e.g. two post boxes cannot both instance the \textit{same} instantiation of the property-type ‘redness’—so all of the property instantiations possessed by \( a \) must be different than those possessed by \( b \), and one of these sets of property tokens must, by the definition of “experience,” be a set of qualia: thus (1) implies (2). Further, by (1), \( a \) must be an external perceptual object and the other individual, \( b \), a bundle of qualia; assuming that all property tokens must inhere in some individual, some qualifier, non-identical with \( a \), must exist which provides the ‘ground’ for \( b \); thus (1) implies (3).

Now assume (2) is true. Then quale \( F \) is non-identical with any property of external object \( a \); it must therefore be a property of some other individual; this individual, by definition, is a qualifier (since it instances a quale): (2) therefore implies (3). Further, (2) means that at least some qualia are not properties of external objects, and presumably these qualia form part of certain experiences; by Leibniz’s law, such experiences cannot be identical with any external perceptual object, since they involve a property token not possessed by any external object; hence (1) follows from (2).

Finally, assume (3) is true; then there must exist some qualifier which is not an external perceptual object. By definition, this qualifier must instance qualia, and so (2) is true. As we have already seen, (2) implies (1), so, by transitivity, (3) must also imply (1).

So, what we are interested in establishing in this chapter and the next is that at least one of (1), (2) or (3) is the case. If any of the theses can be established, then the truth of the other two claims follows more or less automatically. What arguments can we give for the positions (1) to (3)? The twentieth-century literature shows a tendency to conflate these three conclusions into arguments for ‘sense-data’ (or raw feels, or sensa, or ‘qualia’ ambiguously under-

\cite{9} These two claims can be construed as structurally parallel: the first says that there cannot be (actual) redness without something red; the second that there cannot (though “frequently is not” seems adequate for the arguments made below) be phenomenal redness without some red experience.
stood, and so on), and also to collapse together various somewhat different arguments for this ambiguous conclusion. However, since the three theses mutually entail each other this ambiguity does not matter as much as it otherwise would, and on analysis at least three distinct types of argument in this domain can be isolated and described:

a) Arguments from givenness;

b) Arguments from the causal facts of perception;

c) Arguments from perceiver relativity, including the possible inexistence of perceived objects.

I will show here that arguments of type a) are either unsound or circular and dependent upon the results of arguments b) and c); but that b) is suggestive; and that c) is conclusive: that is, some arguments from perceiver relativity rely upon very plausible premises and are demonstrably valid, and so propositions (1) to (3) are (most probably) all true (it actually turns out to be proposition (2) which can be most successfully proven directly). That is, phenomenal properties are (at least sometimes) to be distinguished from external properties and so the experiences they go to make up are distinct from external objects, while the individuals in which

\[10\] The work of Wilfrid Sellars illustrates a fourth type of argument for the existence of qualia and qualifers. Sellars, a scientific realist, attempted to justify the existence of “sensa” by showing that they are necessary for a fully adequate theory of perception within a unified scientific description and explanation of the world. That is, sensa might have a role analogous to that of dark matter today or, at one time, positrons. Sellars argues, contra Quine, that the explanation of certain sorts of human behaviour—for example, the verbal expression of perceptual propositional attitudes such as “there is a red rectangle over there”—requires sensation terms. Perceptual propositional attitudes are elements in an explanatory theory, but they must themselves be explained, and this is done by a person being “under the visual impression that (visually taking it to be the case that) there is (or of there being) a red and rectangular physical object in front of one” (1967a, 14). Sellars insists that sensations are necessary, not to explain discrimination behaviour, but to explain certain features of human language behaviour.

Sellars had various arguments for this position, but they are deeply embedded within his own carefully worked-out world-view, and it would take us too far afield to examine them here. In brief, he held that in the “scientific” (as opposed to “manifest”) image, persons must be seen as bundles of particulars. This means that persons are not logical subjects—in particular not subjects of adverbial construals like “to sense red-ly” or other accounts which treat sensations as properties of persons as whole. The scientific image consists of a unified, complete, true description of the universe that uses only predicates that have individual scientifically basic entities as subjects. This leaves open only the possibility that the sensory experiences of persons are particulars; that is, sensa.
those properties inhere are likewise distinct from the external objects of perception. Phenomenal envelopes are different ‘entities’ from real envelopes; the phenomenal property tokens that constitute phenomenal envelopes are different property tokens than those instanced by the real envelope; and the individuals in which these phenomenal property tokens inhere are not envelopes.

2. ARGUMENTS FROM GIVENNESS

This class of argument contains two related types: the argument from differential certainty, and the argument from the mediateness of perception. I shall deal with them separately, in turn.

The argument from differential certainty was primarily defended (in this century) in the writings of C.I. Lewis (1929, 122 ff.), A.J. Ayer (1940, 19 ff.; 1956, 95 ff.) and H.H. Price (1950, Ch. 2), and was roundly attacked by R.J. Hirst (1959, 32–37). It argues most directly for the thesis (1) above, that distinguishing external objects from experiences, and it has the following structure:

i) In sensing, say, a tomato, one thing one cannot doubt is that there is a certain visual experience\textsuperscript{11} indubitably present.

ii) In a single act of sensing what we take to be a real tomato may be a fake, a reflection, or an hallucination.

iii) Hence, in a single act of sensing no one external thing is indubitably present (from premise ii).

iv) What is indubitably present cannot be identical with what is not indubitably present.

v) Therefore, the experience is not identical with the perceived object (proposition (1) above).

Here, for example, is the way Price first presents the argument:

When I see a tomato there is much that I can doubt. I can doubt whether it is a tomato that I am seeing, and not a cleverly painted piece of wax. I can doubt whether there is

\textsuperscript{11} I.e., here and elsewhere, experience in (at least) my semi-technical sense.
any material thing there at all. Perhaps what I took for a tomato was really a reflection; perhaps I am even the victim of some hallucination. One thing however I cannot doubt: that there exists a red patch of a round and somewhat bulgy shape, standing out from a background of other colour-patches, and having a certain visual depth, and that this whole field of colour is directly present to my consciousness. (1950, 3)

Historically, attacks upon this argument have often treated it as specifically a defence of the doctrine of the ghost qualifier: Hirst, for example, expends the bulk of his critical ammunition rebutting the claim that a red ‘sense-datum,’ say, is “a red existent independent of us and our awareness of it” (1959, 36). However this is not a thesis we are here concerned to defend. More relevantly for our purposes, we can show that the argument is unsound even when construed, less ambitiously, as an argument merely for proposition (1): this is because premise iv) is false—it is not a genuine case of the principle of contradiction.

The principle of contradiction, in one of its forms, states that something \( x \) cannot be both \( Fx \) and not \( Fx \). However “indubitably present,” unlike, say “green” or “square,” need not be analysed as simply \( F \): more plausibly, the simple property in question is “being present”\(^{12}\) and whether it is “indubitable” or not depends on what we might call the modality of the proposition. Thus premise iv) becomes:

\[
iv') \text{ What is indubitably } P \text{ cannot be identical with what is not indubitably } P.\]

But this is not true (even when relativised to a single knower): something might well have a property which is indubitable to that knower under one description or mode of presentation but not under another. For example, I might have no doubt that Mark Twain is the author of *The Adventures of Huckleberry Finn*, but seriously doubt that the same predicate was true of Samuel Clemens; yet Mark Twain is nevertheless identical with Samuel Clemens. Or, taking “indubitable” more radically—more closely to the kind of indubitability involved with experi-

---

\(^{12}\) Whatever exactly *that* means; I won’t worry about this here, but will have some relevant things to say in Chapter Four.

\(^{13}\) This, as Hirst points out, is closely analogous to “What is certainly present cannot be identical with what is possibly present”; in this very similar case, we have no doubt that “certainly present” is best rendered as a modally qualified simple property, rather than as a simple property itself.
ence\textsuperscript{14}—I might be unable to query the truth of the \textit{cogito}, but have no idea what “I”—that thing that undoubtedly exists—really am. Suppose for a moment that my self is identical with my brain; I might have no doubt that I am a thinking thing, but nevertheless be suspicious (as indeed Descartes was) of the claim that my brain is a thinking thing even though \textit{ex hypothesi} they are the same thing.\textsuperscript{15}

So, we must abandon premise iv). Without premise iv), the argument from differential certainty is consistent with the claim that the fact of the presence of an experience of a red tomato is indubitable, while the simultaneous existence of an actual red tomato may be doubted, but that the two entities might nevertheless be the same thing. That is, it fails to establish its conclusion.

A separate strand of argumentation can often be discerned intertwined with claims about the indubitability or certainty of ‘sense-data’\textsuperscript{16}: this involves the claim that ‘sense-data’ are \textit{immediately} or \textit{directly} ‘present to the mind,’ and are on \textit{this} basis to be distinguished from external perceived objects. For example, Price continues the passage quoted just above by saying that, not only is the existence of the bulgy red experience indubitable, but, by contrast with the real

\textsuperscript{14} The notion of “indubitability” is a rather slippery one, which I discuss in detail in Chapter Five. I don’t think much hangs on the shades of meaning of the term for the purposes of attacking the third premise; however, it is possible to use discussion of the concept to attack the \textit{first} premise. Hirst, for example, complains that this “immediacy assumption” requires that any unclarity in perception must be attributed to the sense-datum and not the sensing. “We do not sense indistinctly a red color-patch; we sense clearly an indistinct red color-patch” (1959, 30). This might be found problematic—Hirst certainly thinks it is; however I shall postpone discussion of this interesting point to Chapter Four.

\textsuperscript{15} Hirst paraphrases “indubitability” as “certainty” (after which, he asserts, “the plausibility [of premise iii] vanishes”), and gives as his counter-example: “… [I]t may be certain that there is a car in the garage but uncertain whether it is a Morris or an Austin or some other make; but that does not mean it cannot be an Austin” (1959, 34).

\textsuperscript{16} The discerning reader may be worrying by now that ‘sense data’ are being treated implicitly, and (we shall see) illicitly, as \textit{facts}, or at least as being fact-like, rather than as either particulars or property tokens. Thus, following the principals in this debate, we are allowing ourselves to speak as if ‘sense-data’ \textit{themselves} (rather than propositions about ‘sense-data’) can be “certain” or “indubitable.” I have deferred detailed discussion of this issue for Chapters Four to Six (and especially Chapter Six, section six), but for now this is part of the reason for the scare quotes around ‘sense-data.’
tomato, “this whole field of colour is directly presented to my consciousness” (1950, 3). We can abstract the following argument from the mediateness of perception:

i) In sensing, say, a tomato, we are directly acquainted with some experience.

ii) In a single act of sensing, our perception of the tomato is mediated and indirect.

iii) What is immediate cannot also be mediate.

iv) Therefore the experience is not identical with the perceived object (proposition (1) above).

The problem with this argument is not that premise iii) is false, but that the argument begs the question. Someone who is unpersuaded of the truth of the conclusion is very unlikely to agree at the outset that some “directly present” experience always accompanies perception, while the actual perceptual object is only mediate perceived; by contrast, they will typically hold that only one “perceptual object” is present, that that object and all its perceived properties are “external,” and that our perception is uniformly either “direct” (as no other ‘object’ intervenes)—in which case, the truth of both premises i) and ii) will be denied—or “indirect” (as our perceptual states are the result of a causal chain originating in the perceptual object)—in which case, the truth of the first premise is called into question.17

This alone is not to say, of course, that premises i) and ii) are false—just that their truth is unlikely to be admitted by someone who does not already agree with the conclusion. The success of the argument, then, depends upon independently establishing its premises … and the best way of doing this seems to be by using some of the reasoning involved in the following arguments, those from the causal facts of perception and perceiver-relativity. The argument from the mediateness of perception, in other words, cannot successfully stand alone.

3. ARGUMENTS FROM THE CAUSAL NATURE OF PERCEPTION

A second common variety of argument for the distinction between ‘sense-data’ and external perceived objects (and/or their properties) was based upon the causal nature of perception. I

---

17 I discuss the notions of directness and immediacy in some detail in Chapter Four.
shall argue that this kind of argument can be highly plausible but is still not quite adequate to defeat certain determined sorts of opposition to its conclusions.

John C. Eccles (1965), like many before him, interpreted the causal facts of perception as supporting the contention that there are ‘sense-data.’

… [W]ith perception the sequence of events is that some stimulus to a sense-organ causes the discharge of impulses along afferent nerve-fibres to the brain, which, after various synaptic relays, eventually evoke specific spatio-temporal patterns of impulses in the neuronal network of the cerebral cortex. The transmission from sense organ to cerebral cortex is by a coded pattern of nerve impulses that is quite unlike the original stimulus to that organ, and the spatio-temporal pattern of neuronal activity that is evoked in the cerebral cortex would be again different. Yet as a consequence of these cerebral patterns of activity, I experience sensations … which in my private perceptual world are ‘projected’ to somewhere outside the cortex. (17–18)

However, despite the fact that the chain-like nature of perception is (or at least, was) often adduced as a reason to suppose that something like distinct ‘sense-data’ exist, it is not always clear just how the conclusion is supposed to follow from this evidence. One of the most perceptive and careful critics of the sense-datum literature, James Cornman, expanded the argument in the following way (1971, 212ff.):

i) In veridical perceptions, a perceptual experience of an object is the effect of an external object stimulating a sense organ.

ii) If the perceptual experience of an object is an effect of an external object stimulating a sense organ, then what is perceptually experienced in the perception is also an effect of the stimulation of the sense organ.

---

18 Eccles follows, for example, Russell’s Problems of Philosophy (1912). Ayer, too, often alludes to the causal nature of perception in his discussions of sense-data: e.g. 1940, 9–11; 1956, 91–95; 1973, 74, 82 ff. Frank Jackson’s chapter on “Colour and Science” (1977, chapter 5) also uses a version of this argument to try to show that colours are not properties of external material objects. He argues that, since only “scientific properties” have causal effects upon our brains, these are the only properties we have “reason to believe” that external material things possess; he then states that colours (and the other secondary properties) are non-identical with scientific properties; the conclusion is that colours (etc.) cannot be properties of external material objects. The main difficulty, of course, is with antecedently defending the premise that colours cannot be identified with scientific properties.
Chapter 2: “Arguments for ‘Sense-Data’.”

iii) No object is identical with one of its effects.

iv) Therefore in veridical perception, what is experienced is not identical with a physical object that stimulates a sense organ.

v) In veridical perceptions, either physical objects that stimulate sense organs or ‘sense-data’ are experienced.

vi) Therefore in veridical (and then surely all other) perceptions, ‘sense-data’ are experienced.

The nature of some of Cornman’s comments and objections suggest that he was interested, once again, mostly in the temptingly radical target of the thesis of the ghost qualifier; the argument can, however, be read as directed towards a more plausible thesis—in particular, positions (1) and (3) described above, i.e. that external perceptual objects are neither experiences nor qualifiers. Thus, let us take Cornman’s “what is perceptually experienced” as being an experience in the sense defined above (while “a perceptual experience” should be understood as picking out a perception, again in the sense defined in Chapter One); then the sub-conclusion that makes up step iv) of the argument is to the effect that experiences are not identical with the physical objects that stimulate sense organs. The first four steps of Cornman’s argument can be rewritten, then, as follows:

i’) In veridical perceptions [process], a perception [product] of an object is the effect of an external object stimulating a sense organ.

ii’) If the perception [product] of an object is an effect of an external object stimulating a sense organ, then the corresponding experience [i.e. bundle of qualia] is also an effect of the stimulation of the sense organ.

iii’) No object is identical with one of its effects.

iv’) Therefore in veridical perception [process], the experience [qualia-bundle] is not identical with a physical object that stimulates a sense organ.

---

19 Cornman actually prefers to speak of ‘sensa,’ but for the sake of consistency I have silently converted this term into the generic and ambiguous term ‘sense-data’: I do not think anything about the arguments depends upon this change. Apart from this change, the argument is given in Cornman’s own words.

20 Rather than a sensation, since Cornman insists on veridical perceptual experience.
Steps v) and vi) can be construed as designed to establish that such experiences are underwritten by (are made up of properties of) qualifiers (‘sense-data’ here) which are distinct from external perceived objects and which actually do exist: the argument then, if successful, would establish both (1) and a rather expanded version of (3) (all perceptual experiences are ‘of’ qualifiers which are non-identical with external perceived objects).

Premises i) and iii) in this argument are, I shall suppose, beyond reasonable dispute, and the argument is structurally valid. Cornman, however, rejects the argument on the basis of the claim that premises ii) or v) may be false: he argues that either physical external objects can be experienced—in which case there is no reason to accept ii)—or no external objects can be experienced, in which case the possibility is still open that nothing is experienced at all and so v) is false.21 The first of these options is (on the present construal) essentially the claim that (properties of) the external, physical perceived objects are the experiences—they are “what is perceptually experienced”; then, though the perceiving of those objects would indeed be one of their effects (and thus distinct from them) it would not be the case that what is experienced was also an effect of the perceptual object … and so premise ii) would be false (and iv)–vi) would not follow). Further, even if the perceived objects are not identical with experiences, Cornman suggests, it might be that nothing at all is an experience—that such things do not exist. For example, “it is plausible to claim that in veridical perception external physical objects are perceived but no object is experienced in this sense, because the perceptual experience that occurs when perceiving consists in sensing in some determinate way rather than sensing some object” (216–217).

Presented in terms of experiences (i.e. sensations construed merely as qualia-bundles) as they are here, neither of these two options, I think, is as plausible as when implicitly put forward against ghost qualifiers—that is, they have much less prima facie force when levelled against

21 The move from “physical objects are not what is experienced” to “therefore possibly nothing is experienced” depends upon Cornman’s particular analysis of “to experience $x$” and is not, in my view, altogether plausible: I shall allow, however, that “possibly nothing is experienced” is a claim that might be thought to be independently justified, and so is worth consideration.
phenomenal properties than when levelled against phenomenal individuals. But this is not to say that the versions which strike against phenomenal properties are universally disbelieved: Fred Dretske, for example, has recently argued that ‘qualia’ (by which, I believe, he means what in my vocabulary I would discriminate into qualia, qualifiers and experiences) are identical with external perceived objects and/or their properties (1995), and Michael Tye has put forward a modified adverbial theory of the mind that thinks of itself as allowing no place for such ‘mental objects’ as experiences (1989). Cornman’s objections, then, must be addressed if arguments from the causal nature of perception are to be generally persuasive.

The point that springs most forcefully to mind in considering how to counteract these two kinds of possibility is that the distinction they rely upon, that between “perceptual experiences”—that is, perceptions (here construed as products rather than processes)—and experiences (in my sense), is often implausible. What would it be, one wants to say, to have a normal visual perception [qua product] of, say, a tree without that perception being itself an experience—the ‘visual image’ of a tree? Surely, one wants to insist, visual perception [qua a sequence of products over time, rather than the process itself—the temporally extended “perceptual experience”] exactly consists in a seamless sequence of imagistic experiences. As I look around my office, the image (experience) of my computer monitor is succeeded by that of my telephone, then the view outside my window, some bookcases, and so on; auditory perceptual experience is precisely a sequence of auditory experiences, such as words or chords; olfactory perceptual experience is made up out of smells; and so on. If this is true, then there is normally no room for distinguishing between “perceptual experience” and experiences (except insofar as the latter are (possibly arbitrary) divisions within the flow of the former). Experiences do exist and are the causal effect of (and hence non-identical with) the external objects that are the causes of perception.

However this response will not do; though it may be the expression of a deeply felt intuition, it is nevertheless merely an expression of intuition, and will not convince those who do

---

I address Dretske’s theory of qualia directly in Chapter Eleven.
not share such intuitions. One who wishes to, so to speak, externalise experience can self-
consistently respond by denying that what Cornman is calling "perceptual experience" is itself
made up of a sequence of experiences, and instead insist that it is rather a sequence of, say,
judgements that, or perhaps non-phenomenal representations that, some external object is a
qualifier—is that object which instances the property tokens that together make up the experi-
ence. Alternatively, one who wishes to eliminate experience will refuse to talk in terms of ex-
periences at all, and limit themselves to external perceived objects, non-phenomenal represen-
tations of those objects, perhaps (or possibly just the process of becoming informed about such
objects, without any end-product 'representation'), and, sometimes, judgements to the effect
that those objects have such-and-such properties.

It seems clear, then, that the causal argument, as it stands, need not move the externalist or
the eliminativist with regards to experiences (or alternatively qualifers), because its premises
assume the falsity of both these positions. There are, however, two more forms of the argu-
ment from the causal nature of perception that should be considered before we abandon this
family of arguments. The first uses the causal facts of perception to argue that experiences
and/or individual qualia vary independently of the external perceived objects, with changes in
the medium and the observer, and in fact seem to covary exactly with states of the brain. For
example:

…[T]he states of the intervening medium and of the sense-organs and afferent nerves
only make a difference to the sense-datum in so far as they make a difference to proc-
esses in that: otherwise they are irrelevant. And the state of the external object itself is
only relevant in so far as it indirectly affects the brain, while in hallucination the exter-
nal object is dispensed with altogether. … It even begins to seem conceivable that they
[sense-data] are themselves cerebral events qualified in a certain way. (Price 1950, 30)

We might set this out as follows:

i) Experiences (in my semi-technical sense) vary independently of the external perceived
   objects.

ii) Therefore experiences are not identical with external perceived objects.

iii) Experiences covary precisely with properties of brain states.
iv) Therefore qualifers are probably brain states.

This argument covers similar ground to the successful argument from perceptual relativity considered below, but in this incarnation it is significantly weaker. Premise i) contains the assumption that experiences do indeed exist, and might also be denied by someone who holds that it is not experiences that are observer-dependent, but something like “representations of” or “judgements about” those experiences. Premise iii), though many today might find it plausible, is empirically under-supported: the science of correlating *qualia* (rather than, say, general perceptual processing) with brain states is—if it can even be said to exist at all at this point—still a very young and inexact one. And conclusion iv) is only fairly weakly supported by its premise: covariation, in general, is far from adequate proof of numerical identity.

The last variety of causal chain argument is the “time-gap” argument, which was probably first expressed in recent times by Bertrand Russell (1948). Russell argued that because light travels at a finite speed, certain far distant objects could have ceased to exist before information about them reached one’s eyes; therefore, what we see cannot be identified with the physical object itself:

… [T]hough you see the sun now, the physical object to be inferred from your seeing existed eight minutes ago; if, in the intervening minutes, the sun had gone out, you would still be seeing exactly what you are seeing. We cannot therefore identify the physical sun with what we see. (204)

Ayer expanded slightly on this (1956, 94), giving us, in Cornman’s view, the following argument (Cornman 1971, 219):

i) If something is perceived, then it is perceived during the time a veridical perceptual experience of it occurs.

ii) Any veridical perceptual experience of an external physical object occurs after the stimulus energy transmitted from the object first affects a sense organ.

iii) Therefore if an external physical object is perceived, then it is perceived after the stimulus energy from it first affects a sense organ.

---

23 It can also be found in Brain 1951.
iv) If an external physical object is perceived after the stimulus energy from it first affects a sense organ, then it is possible it does not exist at the time it is perceived.

v) It is impossible that something is perceived at a time it does not exist.

vi) Therefore, no external physical objects are perceived.

vii) If no external physical objects are perceived, then sense-data are perceived.24

viii) Therefore, sense-data are perceived.

As we did before, let us parse Cornman’s “sense-data”25 as “qualifiers which are not external perceived objects” and “perceptual experience” as “perception (qua product rather than process).”

Cornman rejects this argument on the basis of the falsity of v), suggesting that “it is quite reasonable to claim that some stars are perceived when they no longer persist” (221). However, this premise is the point at which Cornman’s reading of the argument goes badly off-track: the point is that it is the “perceptual experience,” rather than the perceptual object, that cannot be experienced (rather than perceived, properly speaking) when it no longer persists. Here is how Ayer originally stated the argument:

…[S]ince in every case in which the light has had an appreciable distance to travel it is possible that the object which we think we are seeing has gone out of existence in the interval, we cannot ever identify it with what we see: for our present experience will be the same, whether this object still exists or not. (1956, 94)

That is “our present experience” can occur without the existence of the actual external perceived object (such as a star), and so the two are not to be identified. The argument, therefore, is better laid out in the following way:

i) If something is perceived, then it is perceived during the time a veridical perceptual experience of it occurs.

ii) Any veridical perceptual experience of an external physical object occurs after the stimu-

---

24 The use of the word “perceive” is not one I would endorse here: I think external objects like stars are properly what is perceived, and qualia and experiences are not perceived but, perhaps, experienced. However I shall shortly recast the argument, and in doing so I will let this problem quietly drop out.

25 Actually, “sensa.”
lus energy transmitted from the object first affects a sense organ.

iii) Therefore if an external physical object is perceived, then it is perceived after the stimulus energy from it first affects a sense organ.

iv) If an external physical object is perceived after the stimulus energy from it first affects a sense organ, then it is possible it does not exist at the time it is perceived.

v) It is impossible that the perceptual experience does not exist at the time it is experienced.

vi) Therefore, experiences are distinct from external physical objects (proposition (1) above).

vii) If no external physical objects are what is experienced, then ‘sense-data’—some other kind of qualifiers—are experienced.

viii) Therefore in veridical (and then surely all other) perceptual experiences, qualifiers that are not external physical objects are what is experienced.

That is, basically, at some time \( t \) the perceived objects may not exist but the experience does; they are therefore not the same thing. This argument is much stronger than the version Cornman sets out and critiques, and is probably the strongest form of argument from the causal nature of perception. The argument is valid and premises i) to iv) are plausible. It still has a weak spot, however, in premise v) (despite our improvement over Cornman’s version): once again, someone who resolutely denies the truth of conclusions vi) and viii) is likely to endorse intuitions that render the truth of v) implausible to them. That is, as with the argument from the distinction between causes and effects above, it is open to the objector to differentiate between the moment of the act of perception—Cornman’s “perceptual experience”—and the experience in our sense, and to insist that the latter is either identical with some structured set of properties of the external perceptual object, or is non-existent. In either case, premise v) will be accounted false.

Nevertheless, despite their flaws, it seems to me that arguments from the causal nature of perception have a great deal of plausibility. Once “sense-datum” is disambiguated and stripped of its ghostly pretensions, the arguments capture what it seems reasonable to think of as a common sense view of the matter: external objects cause (and are distinct from) perceptual states, and among those states (roughly speaking) are experiences—sequences of visual images,
sounds, smells, and so on. To deny this, it seems, is to be in the grip of another, less familiarly intuitive, theory—one which either denies the existence of images and smells (at least as phenomenal entities) or which identifies such bundles of phenomenal properties with properties of external objects, and thus, for example, admits (on the face of it, bizarrely) that the visual image of a tree or a star—the relevant experience—might cease to exist before the “moment of perception” of that object takes place.

These arguments, then, will not convince those who hold non-commonsensical theories of experience—and, by the same token, do not prove such opponents wrong—but they do press the issue in such a way as to make it reasonable to suppose that the burden of proof is upon the enemies of internal experience. The next set of arguments will do even more than that.
Chapter 3: Qualia and the Argument from Perceiver Relativity

The third and final class of arguments for ‘sense-data’—arguments from perceiver relativity—comes in two, closely related, versions: what is usually called the argument from illusion, and the argument from hallucination. I shall consider them in order: in fact, it will turn out, despite the care that was often taken to distinguish between the two forms in the literature, that the latter is virtually a subset of the moves in the former. I shall argue that these arguments do in fact establish propositions (1) to (3) of Chapter Two: they show that qualia must, at least sometimes, be distinct from properties of external, perceived objects (and from this it follows that experiences in my technical sense, and the relevant qualifiers, are distinct individuals from the external objects of perception). Thus, the green one sees when looking at an elm tree in summer—that very property instantiation, since these are the property tokens in question—is not a property of the tree.

From this in turn I shall argue that the phenomenal property tokens in question must inhere in something else, and that the most plausible candidates for these qualifiers, in humans at least, are states of the central nervous system. Since brain states, clearly, are not actually green, this suggests that we must think of phenomenal greenness in a different way than actual greenness: to be phenomenally green—to be a green quale—it turns out, is not to (e.g.) have certain surface reflectance properties, but is (we might say, in a preliminary fashion) to feel or experience green—to be a sensation of a certain phenomenal type.¹ In the following three chapters I shall explore the epistemic consequences of a view like this.

¹ I will suggest below that one can, if it feels more comfortable, think of this as “sensing greenly.” However I think that the adverbial locution is rather looser than the terminology I have chosen (which may well be the reason for its apparent comfort factor): in particular, the adverbial phrase seems to remain neutral on the phenomenal-ity of the sensing—the fact that the property in question is that very one we experience when looking at a tree—which is something I shall wish to insist on.
1. THE ARGUMENT FROM ILLUSION

The argument traditionally called the argument from illusion but re-labelled more accurately the argument from perceptual relativity by, among others, Hirst (1959) and Cornman (1971), has as its basic form the premise that ‘sense-data’ vary in ways that the external perceived objects do not; that therefore ‘sense-data’ cannot be identical with (the manifest properties of) those external objects; and the conclusion that the qualifers and the external objects must be two different things. In addition this argument, as with the others we consider here, can have as a second conclusion the claim that both qualia and qualifers of some sort exist, since something must exist which is “what we see when an object looks quite different from what it really is” (Hirst 1959, 45).

There are a variety of forms the first premise can take. A common one relies on the claim that sense-data vary between individuals when they are looking at or otherwise perceiving the same object. It is also possible to make a very similar argument using intra-personal differences in experience over time when it is plausible to assert that the external object itself has not changed in the relevant respects (for example, changes to the size and shape of objects when the observer's perspective is varied). Alternatively one can appeal to the sensory experiences of

---

2 This, for example, is Moore’s tactic, holding up an envelope before his lecture audience: “Though we all did (as we should say) see the same envelope, no two of us, in all probability, saw exactly the same sense-data” (Moore 1953, 30). Probably, each person saw a slightly different shade of white (depending on the light), a slightly different shape (depending on their perspective), a colour patch of a slightly different size (depending on their distance from the envelope) and so on. Even if two people by chance happened to see exactly the same figure, we should not say we knew that they had; whereas, we can say we know that everyone in the room saw the same envelope. As Moore points out, had they seen him commit murder, they would have had no doubt that it was he, the same man, who should be convicted of it (1953, 33). “Now all this seems to me to show very clearly, that, if we did all see the same envelope, the envelope which we saw was not identical with the sense-data which we saw: the envelope cannot be exactly the same thing as each of the sets of sense-data, which we each of us saw; for these were in all probability each of them slightly different from all the rest, and they cannot, therefore, all be exactly the same thing as the envelope” (Moore 1953, 33).

Notice, by the way, the elision here: the real claim is that the properties of the envelope differ from those making up the different individual experiences, and so the objective envelope (one thing) cannot be the same thing as the various qualifers (many things), but Moore’s way of speaking made it hard for him to notice this.
non-normal perceivers when they observe material objects, such as the perceptual experiences of the colour-blind or drug influenced; or the first premise can rely upon such common perceptual illusions as straight sticks appearing bent when refracted by water, or distant green hills seeming purple from a distance. Finally there is another possible premise, less widely noticed in the literature, which relies on the possibility that other species, such as flies, bats and Martians, might have very different perceptual experiences when observing the same external objects.3

As is well known, the argument from illusion—despite what I would have thought was its initial plausibility—has been widely ridiculed in the recent philosophical past. To closely paraphrase J.L. Austin (1962, 29), for example: “What is wrong, what is even faintly surprising, in the idea of a stick’s being straight but looking bent sometimes? Does anyone suppose that if something is straight, then it jolly well has to look straight at all times and in all circumstances? Obviously no one seriously supposes this. So what mess are we supposed to get into here, what is the difficulty?” Even sense-datum theorists themselves often cast scorn upon “the notorious arguments from illusion, variation, perceptual relativity, and so on and so forth. … [L]et me say straight away that I think these arguments prove nothing. … I believe that the current opposition to sense-data derives in large measure from their unfortunate historical association with these arguments” (Jackson 1977, 107).

Let us, then, try to reconstruct more precisely the argument from perceiver relativity in order to properly assess its strength.4

i) If perceiver $y$ differs from another perceiver $z$ in having a different perceptual perspective, or in seeing $x$ at a different time under different conditions, or in being perceptually

3 William Seager (1991, 150–151, 208)—drawing on the University of Toronto doctoral work of Evan Thompson—notes that pigeon colour vision is subserved by four or five types of colour receptor, as contrasted with the three of human colour vision. Since we know that people lacking colour receptors see no hues, that people lacking one see only two fundamental hues, and that normal humans see three, this implies that pigeons see extra hues compared to humans. (Pigeons don’t have all the advantages, though: they have only 37 taste buds, compared to a human’s 9,000.)

non-normal, or in being under the influence of some perceptual illusion, or in being a member of another species, then it is possible that a physical object $x$ may appear to have two incompatible (phenomenal) properties $F$ and $G$ (such as being both uniformly grey and uniformly red, or being circular and being elliptical at the same time).

ii) Hence, a physical object $x$ may appear to have at the same time two incompatible (phenomenal) properties $F$ and $G$ to two different observers $y$ and $z$.

iii) Since $F$ and $G$ are incompatible, $x$ is not both $F$ and $G$.

iv) If $x$ is $F$ but not $G$, then one of the perceivers is perceptually experiencing a property which is not a property of $x$. Likewise, if $x$ is $G$ but not $F$, then one of the perceivers is perceptually experiencing a property which is not a property of $x$. If $x$ is neither $F$ nor $G$, then both of the perceivers are perceptually experiencing a property which is not a property of $x$.

v) Therefore someone perceptually experiences at least one property which is not a property of $x$.

vi) If someone perceptually experiences a property, then that phenomenal property is tokened: for example, if someone senses a red ball, then normally phenomenal redness is tokened.

vii) Every property token is a property of some individual—there are no unowned properties … all property instantiations are also instanced.

viii) Therefore, either $F$ or $G$ must be properties of some qualifier which is not the external perceptual object $x$ (which gives me propositions (2) and (3) from Chapter Two).

It is also possible to extend this argument to non-deductively suggest the conclusion that all that is experienced are properties of non-external qualifiers:

ix) $F$ and $G$ do not seem to differ with respect to what they are properties of. For example, if you pass from a position where a bowl looks circular to one where it looks elliptical, 

---

5 For simplicity of exposition I will ignore the possibility of intrapersonal relativity of perception and am stipulating that the contrasting perceptions occur at the same time.
there is no sudden change such as might be expected if you changed from sensing an external object to sensing a sense-datum, or *vice versa*. Veridical and delusive sensings may form a continuous, gradually shaded series, and often do not differ in their ‘qualitative,’ ‘intrinsic’ aspects.

x) Therefore, all the perceptual properties of an object are probably properties of qualifiers which are distinct from the external perceived objects.  

In somewhat more informal terms, the deductive portion of the argument can be expressed as follows. Sometimes, external objects of perception present two different, incompatible phenomenal properties (i.e. qualia, as minimalistically defined in Chapter One) to two different perceivers at the same time, or to one perceiver over time. Thus, for example, a ripe strawberry may have the phenomenal property of being red all over to a normal observer, but look completely dark grey to a colour-blind perceiver just inches to the left; yet, one wants to say, the strawberry itself cannot be both completely red and completely grey. Hence at least one quale in such cases is not a property of the perceived objects. In the strawberry case, the fruit looks a different colour to the two different perceivers; that is, there are two different qualia being instantiated, again in exactly the minimalist sense of “qualia” which we pinned down previously. For one person, the strawberry ‘presents’ the phenomenal property of redness; for the other, it ‘presents’ the phenomenal property of greyness. Even if one of these two properties—phenomenal redness, say—inhered in the strawberry, they could not both inhere in the strawberry. Its surface could not both be ‘covered in’ phenomenal redness and phenomenal grey.

Well so, one might ask, what? Among those who accept it as true, this sort of conclusion

---

6 I shall simply take some property to be intrinsic iff it is not extrinsic, where a property is extrinsic “just if to possess *F* is to stand in some relation to other, wholly distinct or non-overlapping, contingent things” (Garrett 1995). (The further subtleties of the intrinsic/extrinsic distinction are discussed in Humberstone 1996 and Lewis 1983b.)

7 This conclusion is apparently endorsed by Moore: “This seems to be the state of things with regard to these sense-data. They seem, in a sense, to have had very little to do with the real envelope…. It seems very probable that none of the colours seen was really a part of the envelope; and that none of the sizes and shapes seen were the size or the shape of the real envelope” (Moore 1953, 38).
(line v) in the argument above) is often taken to be obviously true, and so it might seem that little of interest is likely to follow from it. However, a couple more moves turn this claim into something more exciting. First, we can say that the experiencing of a phenomenal property is the tokening or instantiation of that property (premise vi)); for example, to (consciously) see something red or taste something bitter usually involves the tokening of phenomenal redness (not necessarily actual redness) or phenomenal bitterness. Visual awareness of a red object is, in part, a token instance of phenomenal redness. Second, it is a truism that there are no unowned property tokens—every property instantiation inheres in some individual (premise vii)). For every property token, there is something of which it is a property. Our conclusion now must be that some phenomenal properties are properties of some individual other than external perceptual objects: cases of misperception involve phenomenal property tokens which cannot be properties of the objects of perception, and so must be properties of something else. There must be something which is phenomenally grey when the colour-blind perceiver looks at a ripe strawberry … and it isn’t the strawberry.

It is important to be as clear as possible about what is being argued for here. This argument from perceiver relativity does not establish that, if something looks red but is not, there must be something else which is red. Rather, it establishes that there must be something else which is phenomenally red—something else in which inheres the quale of redness. But this is, in its own way, a radical enough claim; the property we are talking about may not be ‘real redness,’ but it is (as outlined in Chapter One) that property we experience when we look at red, or putatively red, things. There is something, this stage of the argument claims, which is not the external object of perception but which has that very colour property which is tokened when we look at red things. There is something, other than the strawberry, in which inheres the redness as we experience it (which, now, must be distinguished from the actual property of redness itself).

*That is, the argument establishes that tokens of phenomenal redness need not also be tokens of ‘real’ redness.*
2. OBJECTIONS TO THE ARGUMENT

Steps i) to viii) above form a valid argument—a formal proof of this, if one is required, can be found in the appendix at the end of this chapter—and moves ix) and x) tack what is at least prima facie a strong inductive argument onto the end of that. What objections might there be

9 How does this argument differ, if at all, from the argument from illusion as traditionally construed? First, the argument was usually set up as supporting the conclusion that some individual or object of perception, additional to the external object being perceived, must exist. By contrast I argue primarily for the existence of properties distinct from properties of the external object. Further, quite often the argument was presented as moving straight from premise ii) (with iii) implicit) to this conclusion, without very much detail as to the steps in between—almost as if the mere claim that appearances are relative is alone enough to warrant the existence of ‘sense-data.’ Here, for example, is the crux of Ayer’s presentation of the argument from illusion in his Foundations of Empirical Knowledge:

From the premise that the appearance of a bent stick in water may be illusory “it follows that at least one of the visual appearances of the stick is delusive; for it cannot be both crooked and straight. Nevertheless, even in the case where what we see is not the real quality of a material thing, it is supposed that we are still seeing something; and that it is convenient to give this a name. And it is for this purpose that philosophers have recourse to the term ‘sense-datum.’ By using it they are able to give what seems to them a satisfactory answer to the question: What is the object of which we are directly aware, in perception, if it is not part of any material thing?” (1940, 4).

In some of his later work, Ayer was a little more detailed in his laying out of the argument—and rather more critical of its soundness. Here is how he presents it in The Problem of Knowledge (1956, 88):

i) Physical objects may appear to people otherwise than as they really are.

ii) In every case in which an object seems to be perceived there is something which is directly perceived.

iii) That which is directly perceived cannot appear otherwise than it is.

iv) Therefore whenever a physical object appears differently from what it is, something other than it is being directly perceived.

Cornman, likewise (and unlike myself), relies heavily upon the notion of direct perception in his construal of the argument from illusion—though he approaches more closely the argument form I use here since “the most plausible way I find to [make the argument] is to talk not about the relativity of perceptions to perceivers, but rather of the relativity of sensible qualities to perceivers, and to construct the argument so that it proves that objects having such properties are perceiver-dependent” (1971, 196).

Frank Jackson, interestingly, did argue from the necessity of the ownership of properties (which I take to be a central engine of my formulation of the argument) to the falsity of the attribution of pains to phantom limbs (1977, 54 ff.), but he does not apply this to his account of the argument from illusion (107 ff.).

10 These last two steps are sometimes mistaken for the argument from perceptual relativity itself. For example Jonathan Dancy (1995) lays out “the argument from illusion” as follows. One cannot tell from the phenomenology of states of awareness whether they are illusory, or genuine awareness of an external reality; it seems likely,
to these arguments?

a) Premise ii): External Objects Have Incompatible Phenomenal Properties

Premise ii) is probably best attacked by questioning the notion of “incompatible” properties upon which it depends. That is, one might wonder whether external perceptual object $x$ might not, in all cases, actually possess both properties $F$ and $G$. However, thought about in detail, it seems difficult to make this objection even coherent. If $x$ has the property of being wholly grey this apparently logically excludes the property of being wholly red. It is just about logically possible that all of $x$ that is visible to $y$ is grey and that all of $x$ that is visible to $z$ is red, but under certain circumstances, such as where $y$ and $z$ are standing close together (but, perhaps, $z$ is colour-blind), this seems radically implausible. Under these and other circumstances, it would lead to wildly creative consequences for the geometry of material objects which seem utterly therefore, that both kinds of state of awareness have something in common—what Dancy calls their “appearance”—and this also suggests that some extra element is responsible for making one of them a genuine instance of perception (for example, but not necessarily, the right causal connection to the perceptual object). Dancy attacks this argument by simply pointing out that it is far from conclusive: mere introspective identity does not entail that the two states could not be radically different in nature. “It is still possible to suppose that the two states do differ … fundamentally, despite their phenomenal similarities; the argument from illusion merely acts as a reminder that there is at best something awkward about this, and that it would be more attractive to avoid it. But the appeal to introspection is not generally allowed to be conclusive elsewhere, and there seems to be nothing special about the present case to warrant any more respectful attitude to introspection here. So the awkwardness is to be admitted, without being allowed to be decisive.” (422) In light of this, obviously, the best response for one bent on denying the force of the argument from illusion is to insist that there are two quite different sorts of experience which may none the less be indistinguishable to their owner. “The first is the genuine article, and the second, though it is indistinguishable, has nothing in common with the first other than the fact that they are both oasis-experiences” (436).

Of course, this argument of Dancy’s really only points out that moves ix) and x) are not conclusive; a fact we are willing to admit to. As far as steps i) to viii) are concerned, considerations such as this seem at least as supportive as critical: how, for example, is the “indistinguishability” we are concerned with here to be cashed out, if not in terms of phenomenal properties? (See also the discussion of the argument from hallucination, below.)

Or, as it used to be put, the ‘sense-data’ (experiences) of which $F$ and $G$ form parts might both be “parts of the surface” of $x$. Moore raised this objection against himself, in order (at that time) to rebut it (1953, 33ff.).
counter-intuitive: what kind of shape could present one wholly red side and one wholly grey side to two observers who stand on the same side of the thing and each reasonably believe they see, say, all of the southern exposure of a red sphere?\textsuperscript{12}

But perhaps what someone who argues phenomenal redness and greyness are not incompatible means is that the properties of being red and grey are \textit{relational} or \textit{dispositional} or \textit{causal} properties, and that the external perceptual object is both wholly red and wholly grey because it can cause the former sensation in some people and the latter in others.\textsuperscript{13} We might say, then, that it has the power of causing both red sensations and grey sensations, or that it can enter into both the relation of appearing red and that of appearing grey.

We should note, first, that this sort of account is only superficially similar to the familiar relational accounts of secondary qualities, such that we might say an object is red if it presents that appearance to a ‘normal’ observer under ‘normal’ conditions. If the apparent familiarity of this doctrine is attractive, then that is misleading. The whole point of the move we are discussing here is to allow us to say that the object in question is \textit{both} wholly red \textit{and} wholly grey.

\textsuperscript{12} Price (1950, 33ff.) raises a slightly different version of the same problem: how can we compare sense-data with the real surface of material objects unless we know what that surface looks like? How can we assert that some properties are \textit{not} properties of external objects unless we already know what properties external objects really have? He plausibly suggests that we could answer this objection \textit{a priori} by claiming that \textit{material thinghood} itself entails being multiply accessible and spatially unitary (i.e. having a single closed, three-dimensional shape, surface, size, position, etc.). Thus, sense-data cannot all be properties of the surface of material things.

Various attempts were made in the heyday of sense-datum theory to further defend the claim that incompatible sense-data may all be parts of the surfaces of material objects—for example, A.N. Whitehead’s theory of “multiple location” (which distinguishes between those characteristics which characterise something \textit{from a place} and those which characterise it \textit{ simpliciter}, and claims that the same part of the surface of an object can have many different viewpoint-relative sense-data); S. Alexander’s theory of “compound things” (which talks of \textit{compound} material objects, which actually are like all their appearances); and H.A. Prichard’s theory of “appearing” (which analyses appearances into three-term relations or complex facts, such that sense-data are part of the surface of an object appearing to someone to have certain characteristics). However, none of these efforts were ultimately successful.

\textsuperscript{13} This kind of account usually assumes colour (etc.) is a secondary quality, but (with some adjustments) it need not do so. Frank Jackson (1996), for example, defends an account of colour as a primary quality—as the categorical basis of its disposition to look coloured.
(and, say, wholly ultraviolet, since that’s how it appears to bees, and wholly Octorine\(^\text{14}\) since that’s how it would appear to certain possible non-Earthly species, and so on for all possible cases). So appeal to normal conditions, on this view, is totally irrelevant in considering the actual, ‘objective’ colour properties of the strawberry. At best we can elect, chauvinistically, to call objects like the fruit “red,” while recognising all along that it is also, \textit{and just as ‘objectively,’} grey, ultraviolet, Octorine, and so on.

Secondly, and more importantly, this objection misses the point. Even if we were to accept the proposal and multiply the number of properties possessed by objects according to the number of different ways they may appear to different observers, what we would mean by this is that they could bring about different phenomenal properties in different perceivers.\(^\text{15}\) But phenomenal properties are the ways objects appear to us: the \textit{phenomenal} redness of a child’s ball is not its disposition to affect us in certain ways (though that may well be what the colour red \textit{is}—it is what we might call the manifest or occurrent visual property of redness. What is at

\(^{14}\) An imaginary extra colour from Terry Pratchett’s \textit{Discworld} novels.

\(^{15}\) Obviously, the object itself has those physical properties which it has, and we would be sensible to admit that it is these causal powers which typically play a leading role in bringing about our perceptions of the object. Thus, it is undisputed that the object can cause both wholly red and wholly grey perceptions: this is not the issue. The question is something like: \textit{where} is that \textit{phenomenal redness}? Is it ‘on the surface’ of the object, or does it inhere in something else? Perceiver relativity commits us to the latter answer.

\(^{16}\) C. L. Hardin, like many others, distinguishes (e.g. 1988, xxi–xxii) between what he calls “physical colour” and “perceived colour,” where we might identify physical colour with, say, the reflectance properties of objects, while perceived colour is the property exemplified by visual sensation. Here what we are interested in is \textit{perceived} colour, and what is at issue is whether or not it can be identical with physical colour—whether the property exemplified by visual sensation might be a reflectance property of some object. Our answer (like Hardin’s) is that it cannot be so identified.

Interestingly, Hardin insists upon restricting our colour vocabulary to \textit{perceived} colours; thus for him the word “red” refers, not to a “physical property” of external objects, but to a particular kind of quale. He chooses this course partly, I think, because he does not believe there \textit{is} any unique “physical property” which is a plausible candidate for being the referent of the word “red.” In order to avoid having to address private language arguments, and to diminish the appearance of a sceptical resurgence, I have by contrast chosen to assume that colour terminology applies to external, commonly available properties; I have no real \textit{arguments} for this position, however, and my stand is only a provisional one—this dissertation is about qualia and not the metaphysics of colour.
issue is whether this manifest property is the same property as any of the physical properties of the ball, and in particular the physical property of being coloured red. The claim currently being argued for is precisely that physical objects can bring about multiple phenomenal property-tokens for different observers, or under different occasions of observation, and that not all of these phenomenal properties can be identified with properties of the object. Once again, unless we are prepared to say that the object’s surface really is wholly ‘covered in’ grey and wholly saturated in redness, the argument still stands.\footnote{To put it another way, though the capacity to appear red is a very good candidate for a relational property, the perceptual sensation of redness—the manifest way the object appears—is not. Consider the difference between two arguably relational or dispositional properties: \textit{brittleness} and \textit{being taller than}. The perception of colour should properly be modelled on the former kind of relationship, and not the latter. Colour perception is unlike \textit{being taller than} in at least two important ways: it involves a causal connection, and it brings about a physical change in one of the two \textit{relata}. An object objectively “is red,” roughly speaking, if it is prone to causally bring about a certain physical change—a sensation of “redness”—in an appropriate observer under normal viewing conditions. Similarly (though conversely), a physical object is brittle if it responds in a certain way to an appropriate causal influence: that is, it shatters if struck. And the analogue to the quale of redness is the \textit{shattered object}: it is the state that is brought about by the causal stimulus. So, of course, some physical objects are red. But, just as obviously, even on the relational picture of secondary qualities, they are not sensations of redness, nor are they the things in which those sensations inhere.}

\textit{b) Premise vi): Phenomenal Properties are Instantiated in Misperception}

Premises i) to v) of the argument from illusion, I think, must therefore be allowed to be true: if I see a red ball and you, perhaps because you are colour-blind, see a grey one, then, since the ball itself cannot really be \textit{both} red and grey, at least one of us must be experiencing a property
that does not belong to the ball. But what about premise vi)—that there is an instantiation of both properties—redness and greyness—occurrent in the world? Many complaints have in the past been made in this general area—clearly the weakest link in the argument—but in my view they have often been imprecisely set up and so tended to miss the mark, frequently because they take aim at the ghost qualifier thesis. For example, R.J. Hirst promisingly remarks that the argument from perceptual relativity trades upon a “confusion between phenomenology and assertions of existence” (1959, 48), but his subsequent discussion focuses entirely on a common-sense analysis of the causes for variations in perception which does not seem to be to the point, since it conflicts with nothing in the argument (at least as I have set it out). Hirst’s goal, it turns out, is to argue that:

…[C]ommon sense need not admit such data as existents, and can say that they are merely experiences of the percipient; and it is the postulation of existents of such a type, unsensed and without corroborative evidence when unsensed, that is epistemologically objectionable. (65)

However, for our purposes, the existents that are argued for are precisely “experiences of the percipient,” the properties that such experiences involve, and the grounds for those properties, and not the potentially unsensed entities additional to both perceiver and perceptual object that Hirst inveighs against.

---

18 It is also possible at this point to object to premise vii): that there are no unowned property-tokens. However, I believe that, at least in the literature on perception and experience, this tack has never really been tried. (As Frank Jackson once put it, “it is … quite clear that it is essential to the notion of a property that it cannot be instantiated in the absence of a bearer” (1977, 54).) Elizabeth Wolgast (1962) did argue that the best response to the argument from illusion is to insist that qualities can exist without belonging to things. However, by this it turns out that she means only that they are merely “appearances” rather than “qualities possessed by things”—they are not properties of external perceived objects. (She thus distinguishes between the ‘real’ colour of, for example, the sky and distant objects, and ‘apparent colours’: between “seeing blue there” and “seeing something blue there” (466). This contrast she labels as being between the “here-now” senses of quality words and the “predicated” senses of those same words.)

19 “One may get a distorted, distant or indirect view of a thing, instead of a good one, as the result of one’s position at an angle or at a long distance from it, or because of intervening media such as mist or water; or one may fail to see or hear it properly because of defects in the sense organs or nervous system” (1959, 48) and so on.

20 J.L. Austin’s famous attacks, too, appear to work only against the “immaterial object” version of sense-datum
A second variety of mistaken objection to premise vi) is of the sort that parses the claim made there simplistically as being: “if something appears $P$, then some object is $P.”$ Cornman (1971, 199ff.) is a good example of this line of thinking (though he was far from alone in it).

His particular counterexample is the property of ghostliness: that something looks ghostly, he says, does not mean that there are ghosts. But, as this counterexample makes clear if it was not before, Cornman’s paraphrase does not capture the content of premise vi) as we have formulated it here: better would be something like “if something appears $P$, then something appears $P$,” or more informatively, “if the property of appearing $P$ is being manifest, then the property of appearing $P$ must inhere in something.” Here, for instance, is the preferred account of ghostliness consistent with premise vi) as we have actually formulated it: Suppose that stunted tree over there in the mist appears ghostly; then it is certainly true to say that the tree itself is what looks ghostly. However, it is not true that the tree itself is coloured a pale grey, or that it has a hazy outline … it is not true, in this sense, that it actually is ghostly (i.e. resembling an apparition). However, neither is it true to say that these phenomenal properties do not exist at all—they are, after all, the reason why the question about the ghostly tree arises in the first

---

21 The premise has also been attacked in this way by, among others, Austin 1962 (if a barn looks like a church, then does one see a church?); Chisholm 1966 (if a man looks tubercular, does it follow that what I see is tubercular?); and Grice 1961 (if some food looks indigestible, then am I looking at an indigestible sense-datum?).
place. These properties, further, cannot be properties of nothing at all: instead they are properties of, so to speak, the appearance of the tree—they make up the experience which is the visual image of the ghostly tree, and inhere in some qualifier distinct from the tree.22

So what attacks can be made upon premise vi) as it actually needs to be stated for the argument to go through? Clearly, to deny the truth of the premise is to deny that if someone perceptually experiences a property, then that qualitative property is instantiated or tokened. That is, it is, like David Armstrong,

---

22 A typical response in the literature of the time to objections of Cornman’s sort was to construe the relevant verbs of appearing as being used in a certain special way: Chisholm, for example, called it the “phenomenological” use (1966), Price the “literal visual” manner (1964), and Grice the “direct” use (1962). This sort of move typically involves distinguishing between “non-presumptive” statements descriptive of one’s visual field, and statements that go beyond this “ultimate evidence.” Kiteley (1972) gives a sequence of examples of this sort of distinction: an art expert might say that a newly found painting in the style of Vermeer “looks” spurious—a presumptive use of “looks”—and then go on to say how the brush strokes or the signature “look”—now using the term in the way we are searching for: “artless, non-presumptive, directly descriptive of how some object presents itself to his senses” (203). Similarly a plate can “look” round when really it “looks” oval, a lake can “look” cold when it literally “looks” leaden and white-capped, and a sign can “look” octagonal when one expects “that a visual count of its sides would make them eight” (204).

This response is on the right lines, but was flawed by its failure to remove what we might call “existential angst”: as Kiteley points out, “if this half-immersed stick looks bent to me then what I see is bent,” even on an approved reading of “looks,” is still analogous with claims like:

a) If this fruit will soon turn rotten then what this fruit will soon turn is rotten; or
b) If he failed to become angry then what he failed to become is angry.

And these conditionals entail existential claims of the following sort: there exists something that is both what he failed to become and angry—the consequent of the conditional goes beyond the antecedent in that it “purports to mention something that has succumbed to [anger]” (206). Thus, it looks like we must also be committed, on this analysis of the illusion of a bent stick, to something that actually is bent. (Frank Jackson, in fact, argued directly from this sort of analysis of phenomenological uses of “looks” to the existence of non-bodily (yet three-dimensional and spatially located) ‘sense-data’ (1977).)

The proper solution, as I suggest above, is to reject the claim that “if something looks P then I see some object that is P,” and replace it with our, much more trivial, premise vi): “if something looks P, then the property of appearing P is being instantiated”—that is, if the half-immersed stick looks bent to me then what I see (phenomenologically speaking—i.e. what I experience) has the property of (so to speak) looking bent, and all that is existentially entailed is something that looks bent (is a sensation of a bent thing), not that actually is bent.
... to deny that when something physical looks green to somebody, but is not green, or where somebody images something green, then the sensory quality of greenness is present” (Armstrong and Malcolm 1984, 171, my italics).

Notice that this claim is not merely that for some reason, “qualia do not exist”—a general position that is these days frequently to be met with, which may well be in opposition to the contention put forward here, but which does not in itself necessarily address any of the premises in the (valid) argument we are considering—but that premise vi) is specifically incorrect; that on those occasions when we experience greenness but where no external perceptual object is objectively green, there simply is no greenness going on—the “sensory quality of phenomenal greenness” is, in fact, not present. That is, Armstrong apparently means, not just that there is nothing green present (which is obviously true), but that there is nothing phenomenally green present—no green qualia: this, at least if it strikes against our thesis here, it is an attack upon the very existence of experienced colours (tastes, and smells) when they do not correspond to the actual colours (tastes and smells) of perceived objects in the world.

How does Armstrong defend this—I think very radical—claim? He confesses that his major motivation for this stance is that allowing we can “experience greenness” when ex hypothesi there are no green objects there casts us behind the veil of appearance.

Once one has gone this far, it proves difficult to maintain that anything except mental things are green. The greenness of vine-leaves is dismissed as a mere façon de parler. Vine-leaves are ‘green’ because they have the power to create in us mental phenomena which have the actual quality of greenness. (171)

---

23 I shall consider more general oppositions to qualia—such as that of Dennett in “Quining Qualia”—in a later chapter; though, if I am correct in judging the argument from perceptual relativity sound, it is unlikely that critiques or theories which deny the argument’s conclusion are likely to be effective.

24 In other words, if you and I both look at a strawberry, and I am colour-blind, must we admit that two phenomenal properties are tokened in the world, phenomenal redness and phenomenal greyness? Could we somehow say that $F$ and $G$ are not both property tokens—that, perhaps, only one of them is?

25 He does add that “there is no call to treat illusory or imagined sensible qualities, and in particular colour, sound, taste, smell, heat or cold, as actual qualities of actual entities” (174), but is not clear that he has any reasons for this claim additional to his modus tollens—like move from the undesirability of perceptual veils to the non-existence of qualia; and this claim itself is not so much an argument against the one discussed here as simply a denial of its force, possibly given additional but unwarranted power to persuade by a hint of the threat of the
Similarly, in *A Materialist Theory of the Mind* (271 ff.), Armstrong refuses to accept that phenomenal properties can be properties of the brain since then we would be “forced to accept a Representative theory of perception, with all its difficulties, unless, indeed, we accept the still more desperate doctrine of Phenomenalism” (272).

Armstrong’s worry, though rather overstated, does have some basis: the argument from perceptual relativity (at least if steps ix) and x) are accepted) does indeed have, as its consequence, the ‘virtual reality’ brain—one where the objective properties of objects in the external world are not, so to speak, ‘directly experienced,’ but are re-created in mental images of the world. This is not, however, to say that objective properties are not ‘directly perceived,’ nor that the ‘actual’ quality of greenness is a property of qualifiers (say, states of the CNS). The facts of the matter, even after the argument from perceptual relativity has gone through, may, presumably, be something like this: objectively green (or spherical, or warm) objects in the world have the power to cause in us, under normal conditions etc. etc., the perceptual sensation of greenness (or sphericity, or warmness), and this experience is made up of phenomenal properties (qualia) which inhere in some qualifier, probably a particular kind of state of the CNS. If this kind of a story is, for some, too bizarre or worrying to be tolerated, then that is unfortunate … but it surely does not constitute a reductio.

ghost qualifier. My position, of course, is that there is, not just a reason, but a compelling reason to treat “illusory sensible qualities”—which, then, are not actual qualities of actual external entities—as nevertheless properties of something.

Jackson calls this “the most widely canvassed objection to Representationalism, that it makes the external world it posits unknowable” (1977, 141). Jackson notes that Representationalism in no way rules out that the evidence of our senses provides good reason for believing in the external world, and he compares the case with the evidence for the molecular theory of gases, which is strongly inductive and not a matter of logical deduction or ‘direct’ perception of molecules (1977 141 ff.). Here, as mentioned below, I side-step the issue by denying that (I am committed to the view that) we infer the existence of the external world from the existence of mental objects; instead, I claim only that our sensation of the external world involves qualia. I suppose it is also true that these sensations could in principle be massively falsidical yet the qualia remain the same (as long as the relevant brain states remain the same) … but I would have thought that something like this (modulo one’s views on qualia) must be true for any theorist who accepts the very possibility of falsidical sensation.
Thus Armstrong’s worry is ungrounded, at least as it applies to this version of the argument from perceptual relativity. First, it is perfectly possible to avoid—and I have tried to do so—the conclusion that greenness is a mental property and not a property of external objects. That is, after all, the point of distinguishing between phenomenal greenness—which it turns out is, at least sometimes, mental—and the “actual quality of” greenness, which continues to be whatever property it is—standardly, some dispositional property of perceived objects. So really business continues as usual: apples, vine leaves and parrots are ‘really green,’ and some sensory experiences are phenomenally green—they ‘feel’ green.

Second, this version of the argument from perceiver relativity does not commit us to the claim that we infer our perceptual knowledge of the external world from our apprehension of our own qualia. It demonstrates only that our sensation of the external world involves qualia, and these qualia are (at least sometimes) not properties of the objects perceived. It is not (intended to be) an indirect act-object theory of perception, and tells no story at all about intentional content. It is certainly possible to tell a story about the content of our sensory states that makes no mention at all of their phenomenal properties, but instead relies upon, say, their teleological or causal properties. That qualia must be brain properties surely does not automatically falsify these theories—it merely tells us more about the character of our mental representations. Thus, the assertion that qualia are brain properties does not entail that the existence or nature of the external world is merely inferred.

Moreover, a position like the one I attribute to the Armstrong of 1984 (and which may be also attributable to Fred Dretske (1995), J.J.C. Smart (1959) and, in some moods, Daniel Dennett\(^\text{28}\)) seems untenable. Presumably the claim would be something like this. Normally, when

---


\(^{28}\) Dennett argues in the following circle (1994, 130–131). Redness is the dispositional power of certain surfaces of physical objects to produce in us a particular discriminative state. “And what are the properties of these internal states? Here we can … play Locke’s card a second time.” The (mechanical, electrical, etc.) primary properties of these states support various secondary, dispositional, properties, such as the disposition to make verbal judgements. And “[t]he semantics of these statements makes it clear what colors supposedly are: reflective properties
we look at something green the property we see—the phenomenal greenness—is a property of the object itself. Vine leaves and their ilk are phenomenally green, but nothing ‘in the head’ is. However, sometimes when we think we see something green, such as when we hallucinate the vision of an oasis, we are in fact mistaken—there is really nothing green around. In these cases, then, since phenomenal greenness is identical with actual greenness (the class of phenomenal properties is a subset of the class of external, objective properties) nothing is phenomenally green. We may well come to believe that some putative external object is green, but this belief involves no qualia, no phenomenal greenness—it differs only in its causal input-output relations (1984, 172–173) from all our other beliefs, such as, for example, my belief that if there is an interpretation in which a sentence S is true then there is an interpretation in which S is true whose domain is enumerable.

It would follow from this, absurdly, that while perception involves qualia, misperception does not. Suppose I look at a green lawn under normal viewing conditions; the Armstrongian would say, reasonably enough, that I experience a green quale—that I experience phenomenal greenness. Suppose, however, that I now put on some red tinted lenses, and that this (as one would say) makes the lawn look magenta. The lawn is not magenta, it is green, and as we agreed with respect to premise i) it cannot be both all green and all magenta at the same time. The glasses are not magenta, they are red. And for the Armstrongian the lawn and glasses are the only objects in the vicinity which have relevant phenomenal properties—they are the only things possessing what Armstrong calls “sensory qualities.” Therefore, the Armstrongian must

---

29 “I think we should grasp the nettle, and simply deny that, because we know very well what the real properties of things are, despite their sensible appearance, this implies that we are not also under illusion at the same time. It is simply that the former belief is dominant over the latter. We are deceived, although we are not deceived.” (Armstrong 1955, 99)

30 This theorem, of course, was proved by L. Löwenheim in 1915.

31 We should recall at this point that ‘misperception’ can reasonably be taken to involve, not just cases of delusion, but all the possible cases of perceptual relativity outlined above—arguably, that is, most of (if not all) our perceptual experience!
conclude, **nothing** is magenta (phenomenally or otherwise)—there is no phenomenal magenta instantiated at all. Certainly, I believe the lawn looks magenta—perhaps my sensory state somehow has the content that the lawn is magenta—but this is not the point. What’s relevant is that the (phenomenal) **colour** magenta is nowhere to be found; indeed, presumably, since **ex hypothesi**, there is no green (or any other colour) either—my sensation is a magenta one, not green, blue, or any other colour—there are no colour qualia associated with this instance of perception at all. This conclusion is generalisable to any case of misperception; hence, for instance, only veridical visual perception is coloured. I consider this a *reductio*.

Apart from the intrinsic implausibility to this claim itself, it leads on into an, if possible, even less pleasant dilemma: either “qualia,” in the sense it is being used in the models of Armstrong and Dretske, have nothing to do with what mental life **feels like**—in which case we are still owed an explanation of the subjective, phenomenal qualities of consciousness, and are back where we started—or, by contrast with veridical perception, **there is nothing it feels like to misperceive**. In short, either the absence of qualia makes a difference to perceptual experience, or it does not, and either way the denial of premise vi) (even combined with the identification of qualia with objective external properties) is deeply unpalatable.

---

32 There is also the problem of identifying just what belief-analogue particular perceptual experiences are supposed to be: for example, one cannot say that perceiving something red is just like believing that the object is red—which is surely the most promising line—since one may perceive it as being red but not believe that it is red (if it is a white wall under a red light, for instance).

33 Along similar lines, it also follows that when one experiences an orange after-image with one’s eyes closed, or undergoes some sort of drug-induced psychedelic hallucination in a darkened room, **there are no colours of any type instantiated at all**. That is, on Armstrong’s view, all colours are external, and there are no relevant external colours around in these cases. Therefore, there are no colour-tokens at all: no orange, not even “apparent, but not real, objective orange”—the whole **point of** views like Armstrong’s is that colour-triggered beliefs are not themselves phenomenally coloured—just (fallacious) non-phenomenal **beliefs** that orange is being sensed.

34 Apart from these *reductio* of the opposing view, are there any more positive arguments in support of premise vi)? The most central consideration is just the evidence of our senses: whenever we undergo visual experience there is colour and shape; whenever we have aural experiences there is sound; whenever we have olfactory experience there are smells; and all of this is so even when we misperceive. (Can anything be more obvious than that an orange after-image actually **is** phenomenally orange—**it looks orange**?) That is, these phenomenal property-
My judgement, then, is that premises i) to vii) must be allowed to hold. It follows that the conclusion of the argument must also be true: there must be at least some phenomenal properties that are properties of qualifiers which are not external perceived objects. That is, both propositions (2) and (3) above are true, and from them follows proposition (1) about experiences.

c) Steps ix) and x): The Inductive Generalisation

Finally, then, what about steps ix) and x) of the argument from perceptual relativity, which try to extend its conclusion to all qualia, qualifiers and experiences? This sub-argument, unfortunately, is based upon the indiscernibility of “absolutely veridical” and “perceptually relative” perceptual experience, and is therefore susceptible to various counter-arguments (which will be noted in more detail in the section on the argument from hallucination just below). However, though not valid, it does I think have substantial inductive force: it shows that it would be, at least, rather strange if exactly one of the indefinitely large array of possible subtly differentiated visual experiences of an envelope were to be identical with the envelope itself—especially since, as Moore pointed out, we are still allowing that everyone sees the envelope whether or not they see the ‘real’ shape, size and colour (1953, 35).

The best way to firm up the argument for the universal conclusion, I think, would be to attack the conceptual distinction used here between “absolutely veridical” and “perceptually relative”—if it could be shown that the notion of an ‘absolutely veridical’ phenomenal property or tokens are certainly present—and this is all that is needed for vi). It is hard to see what more can be said.

H. H. Price did once suggest a somewhat more sophisticated argument in favour of premise vi) when he wrote that: “When something looks $\phi$ in this literal and visual sense of the word, there is no denying that ‘looks’ comes very close to being a kind of ‘is.’ … When something looks $\phi$, we are provided with an ostensive definition of the concept ‘$\phi$.’ Does it not follow from this that we are being aware of a $\phi$-ish particular, an entity which does actually exemplify or instantiate the characteristic $\phi$?” (1964, 16) However, it is not clear to me what is really added by the swing through the notion of an ostensive definition—especially since Price’s wording re-introduces the risk of committing oneself to the claim that if something looks bent then something is bent … or perhaps, even worse, that what it is to be bent exactly is to look bent.
experience is somehow incoherent, then the universal conclusion goes through (as does the more limited conclusion we have already demonstrated through other means). However, this is an ambitious project that I will not attempt here; I shall rest content with having proved propositions (1) through (3), especially since I think the important conceptual hurdle was the establishing of the existence of any non-external qualia and qualifiers at all.

3. The Argument from Hallucination

At this point, we should, as promised, quickly consider what was labelled in the literature the “argument from hallucination,” since this is another very similar argument for the same conclusion. Essential to these special cases of the argument from illusion is that there is no external object at all that might even pre-theoretically be taken as having the relevant sensed properties. As Hirst puts it:

> In these contexts an hallucination is an experience subjectively indistinguishable from genuine perception, but one in which what a person takes to be a physical object has no public existence or cannot plausibly be identified with any public object. (1959, 37)

The argument from hallucination has traditionally been cashed out in terms of indiscernibility: the whole point of a hallucination, from this perspective, is that “the intrinsic characteristics of the sense-data are exactly the same [as genuine cases of perception], otherwise we should not be deceived, but should by simple inspection be able to distinguish genuine and hallucinatory experiences” (Hirst 1959, 39). Thus, for example, Sir W. R. Brain argued that,

> … if having a hallucination to which no object corresponds is a sensory experience in itself indistinguishable from seeing a real object, this is a strong argument for the view that seeing a real object also involves experiencing a sense-datum which is generated by the brain and is therefore independent of the object. (1959, 10)

This consideration is sometimes reinforced by the fact that in experiences which are only in part hallucinatory there is no intrinsic distinction between the hallucinatory and veridical elements of the experience. Brain quotes from someone who describes his experiences after having taken lysergic acid:

> Then my eyes went to the whitish-gold distempered wall above, where the lamp-light
fell. The wall began to be covered with an incredibly beautiful series of patterns—embossed, drawn, painted, but continuously changing. More colour. Indescribable colour. And all the colours, all the patterns, were in the wall in any case—only we don’t usually see them, for we haven’t eyes to. (Brain 1959, 12)

This has suggested an argument of the following form (slightly modified from Cornman 1971, 203ff.):

i) There is no way to discern from any internal properties of, and relations among, one person’s sensations that some of his hallucinatory sensations are not veridical sensations.

ii) If a sensation of one kind is indistinguishable in kind from sensations of another kind, then the kind of object sensed in the one is indistinguishable from the kind of object sensed in the others.

iii) If two entities are indistinguishable in kind, then they are identical in kind.

iv) Therefore the kind of object sensed in some hallucinatory experiences is the same kind of object sensed in veridical perceptual experiences.

v) In veridical perceptual experience, either external physical objects or non-external qualifies are sensed.

vi) No external physical objects are sensed in completely hallucinatory experiences.

vii) Therefore in some hallucinatory sensations and in veridical perceptual sensations non-external qualifies are sensed.

viii) If non-external qualifies are sensed, then there are non-external qualifies.

ix) Therefore there are non-external qualifies (proposition (3)).

I have taken Cornman’s talk about “kinds of objects” as being about the entities in which the perceived properties inhere (and which are thus sensed (as opposed to perceived)).

Set up in this way, however, the argument can be refuted as it moves through the first three

---

35 For example, Ayer (1956, 90) formulates the argument from hallucination—which he by this time considers stronger than the unadorned argument from illusion, but still rather unpersuasive—as being based upon first-person indiscernibility. In cases of hallucination, “it is argued, there would not have been anything in the character of the experience, considered simply in itself, to differentiate it from one that was not delusive. It is because an experience of this sort is like the experience of seeing a real physical object that hallucinations are possible.”
steps. Given that i) is true, all that follows is that the conditional in ii) deals with \textit{phenomenal} indiscernibility, rather than indiscernibility \textit{simpliciter}, and so iii) becomes “if there is no way to discern from such phenomenal facts whether two things differ in kind, then they are identical in kind” (Cornman 1971, 206) which is surely false. That is, one can argue, with Dancy 1995, Hirst 1959 and Austin 1962, that phenomenal or introspective indiscernibility carries little weight in assessing actual indiscernibility. It does not follow from the fact that I am mistaken or deluded that the two things must be qualitatively indistinguishable, Austin says (1962, 50–51). That I fail to distinguish does not make something indistinguishable: think of wine vintages or conjuring tricks.

In addition, some have made the fairly plausible claim that hallucinations and veridical perceptions are \textit{not} phenomenally identical and “their confusion with external objects … [is to] be explained by the circumstances in which they occur” (Hirst 1959, 42). Austin pointed out that dreams usually feel different than reality; we speak of a “dream-like” quality in art, and so on. Dreaming of being presented to the Pope is not like actually being presented to the Pope. After-images are not like colour patches on walls; a wall seen through blue spectacles is not like a blue wall; a stick refracted in the water is not like an ordinary bent stick (out of the water); and so on (Austin 1962, 47–48). Finally, why should we “expect” two things which are of two different kinds to be qualitatively different? There is often no big qualitative difference, for example, between seeing things and seeing their reflections.\(^{36}\)

In reality, however, the argument from hallucination, properly set out, escapes all these objections: it is more charitably seen, \textit{not} as an argument from indiscernibility, but as the following close relation—indeed, almost subset—of the argument from illusion:

i) During some cases of hallucination, perceptual experiences occur when there is no relevant external perceptual object present at all.

ii) Hence, in some cases of hallucination, someone perceptually experiences at least one

\(^{36}\) It is also common to object that ii) illicitly assumes the existence of some \textit{object of} experience during hallucinations, while further conditionalising the premise to add the antecedent “if there is an object experienced at all” conditionalises the argument all the way through and so the argument becomes practically circular.
property, \( F \), which is not a property of any external perceptual object.

iii) If someone perceptually experiences a property, then that phenomenal property is instantiated or tokened: if someone experiences, say, a red after-image, then phenomenal redness is tokened.

iv) Every property token is a property of some individual: there are no unowned properties; all property instantiations are also instanced.

v) \( F \) must be a property of some qualifier which is not an external perceptual object.

Clearly, this version of the argument from perceptual relativity does not really add anything to what is already contained in the argument from illusion. However, it does make more intuitively pressing the case of phenomenal properties, such as greenness and high pitched humming noises, which could not possibly be properties of external perceived objects, and therefore must be properties of some other qualifier, or not property instantiations at all. Yet, in such cases, \textit{ex hypothesi} there \textit{is} greenness or humming going on—that is just what it \textit{is} for one to be experiencing a green after-image or a buzzing in the ears. So \textit{this} greenness or humming sound must be the property of some non-external qualifier. That is, this version of the argument avoids any worries one might have about premise i) of the argument from illusion, and focuses all attention upon premise vi). Since the argument from illusion is sound, so is the argument from hallucination.

4. THE ADVERBIAL CRITIQUE

I take it then that we now have in our hands at least one strong argument for the position that phenomenal properties—colours as we see them, sounds as we hear them, and so on—are non-external properties, and we have considered and found wanting various possible objections to the premises of this argument. There remains to be considered one important, somewhat more oblique, potential objection to the argument from perceiver relativity. At the time sense-datum theory was reaching its peak in this century—around about the late 1940s—probably the most popular and influential competing view was the “adverbial” alternative.
Crudely, this was the position that ‘sense-data’ do not exist, and that instead the language of “sensations” and “appearances” should be interpreted adverbially: thus, for example, the claim that something looks red to one should be interpreted as something like the claim that one is currently “perceiving redly”; there is no “red appearance,” only a case of perceiving redly. One can experience redly and circularly without there being, in the mind or anywhere else, red circles (as in the case of dreams and hallucinations). I want to rather briefly consider the claims of this influential opponent to sense-data here.

There is an intuitive core to the adverbialist position that, I submit, gives it most of its initial force … but which is really nothing more than a rather obfuscating restatement of the position outlined above: that is, the adverbial thesis at bottom is actually not an attack on the position I am defending here, but rather a (somewhat vague and misleading) reiteration of it.

Adverbialism is best seen, in my view, as an attempt to make a common sense response to the perceived excess of sense-datum theory—as Chi sholm, a leading exponent of the adverbial theory, described his aims:

... [W]hat we want to do is to describe perceiving in that way which is least puzzling philosophically, [and] ... this strange and artificial terminology would seem to be the least misleading. The alternative terminologies entangle us in philosophical questions we can avoid if we talk in terms of sensing. (1957, 124)

This construal of the adverbial thesis treats it as an attack upon ghost qualifers in which inhere the properties we perceive (as opposed to those we sense). That is, it consists in a denial of the claim that there is some ghostly intermediate object in addition to those individuals involved in the physical/neural process of perception (more or less as commonly understood); and it is a denial of the claim that the properties we perceive (objective greenness, objective squareness) are properties of anything but external perceived objects. Neither of these theses is inconsistent with what has been established in the previous sections.

Further, adverbialists typically concede that, for example, sensing redly can occur even in the absence of objective redness—even without red objects in the perceptual field. What is it, then, that makes this sensing redly rather than, say, yellowly, painfully, wistfully or quickly? It is
open to (though not actually required of) my hypothetical adverbialist, at this point, to agree that kinds of sensing are identified by their associated phenomenal sensations—that what makes a kind of sensing *sensing redly* is, at least in part, the phenomenal redness that accompanies such sensings. That is, “redly” might be construed as a *phenomenal* adverbial property, such that there is some particular ‘feel’ to sensing redly, and a different one for sensing yellowly.

Having come this far one wants now, in line with the discussion above, to ask in what individuals this property of phenomenal redness inheres. The die-hard adverbialist will presumably want to claim that phenomenal redness is itself a property of the *process* of perception or sensation; personally, however, I find it more plausible (or accurate) to say that phenomenal redness is a property of some of those *states* whose sequence makes up a process of perception, and in particular, some states of the brain of the perceiver. This has the advantage that phenomenal redness need not be a property of the whole process of perception, from external object—or at least afferent nerve endings—to its conclusion (whatever exactly that is). Phenomenal redness just does not seem to be the kind of property that is possessed by, say, retinal patterns or states of the lateral geniculate nucleus... and we have already shown that at least sometimes it is not possessed by external perceived objects. Further, phenomenal redness does not seem to be the kind of property that can be possessed by a process *per se*—it does not, unlike such potentially adverbial properties as speed or delicacy or intelligence, seem capable of *modifying* a process: this is probably the root of the strangeness to the sound of “sensing redly” that Chisholm admits to above. Moreover, it is not inconsistent with the adverbialist’s position—at least as construed here—to accede to my suggestion and to call some process “sensing redly” if it involves states which have the property of phenomenal redness.

---

This certainly seems to be the view of C.J. Ducasse, an influential early adverbialist, who argued that “‘blue,’ ‘bitter,’ ‘sweet,’ etc., are names not of objects of experience nor of species of objects of experience but of *species of experience itself*” (1942, 232), and that “our sensing blue of a determinate species, i.e., our sensing bluely-in-some-completely-specific-manner, constitutes occurrence of a case of blue. That is, it constitutes *presence at a determinate time* of blue of that determinate shade, and therefore of course, of blue” (234).

An area through which signals from the eyes pass, and in which preliminary processing is done, before the information is moved on to the visual cortex.
Now, that there is a plausible kind of adverbialism which is no objection to any of propositions (1) to (3) above, does not mean that all forms of adverbialism provide no criticism of it. Indeed, in their hurry to rule out ghost qualifers, many protagonists of adverbialism slipped into the kind of language that prima facie conflicts with what we have established above—partly, I think, because they

a) tended to label ghost qualifers with words that suggest what we are calling experiences (such as “sensations,” “looks,” or “experiences”); and

b) were concerned mostly with external, and not phenomenal, properties.

Consider the following quotation from J.J.C. Smart:

To say that an image or sense-datum is green is not to say that the conscious experience of having the image or sense-datum is green. It is to say that it is the sort of image we have when in normal conditions we look at a green apple, for example. Apples and unripe bananas can be green, but not the experience of seeing them. An image or sense-datum can be green in a derivative sense, but this need not cause any worry, because, on the view I am defending, images and sense-data are not constituents of the world, though the processes of having an image or sense-datum are actual processes in the world. The experience of having a green sense-datum is not itself green; it is a process occurring in grey matter. The world contains plumbers, but does not contain the average plumber; it also contains the having of a sense-datum, but does not contain the sense-datum. (1963, 161)

With most of this we can agree: experiences are not green in the same sense that apples and unripe bananas are; they are phenomenally green—sensations of greenness. Further, we can agree that there is nothing more involved in perception than external material objects and certain psychological processes, including the “process of having an image.” However actual processes are made up of sequences of actual states or events; these states, Smart must concede,

---

39 Ducasse expresses this sentiment admirably: “…what I mean when I refer to blue as content of my awareness of my awareness of blue is that my awareness is at the moment of the determinate sort called ‘blue’, and not that it has, like lapis lazuli, the property of being blue; for when I assert, of lapis lazuli, that it is blue, what I mean is that it is such that, whenever I turn my eyes upon it in daylight, it causes me to experience something called ‘blue’; whereas I mean nothing like this when I say, of my awareness, that at a given moment it is of the particular sort called ‘blue’” (1942, 236).
Chapter 3: “Qualia and the Argument from Perceiver Relativity.”

do exist (unlike the average plumber) … and some of these states are indeed what Smart calls “derivatively” green—that is, phenomenally green (or at least Smart has not ruled this possibility out). It is misleading, then, for Smart to conclude that ‘sense-data’ are mythical; certainly, extra, non-physical objects are mythical, but this does not mean that things which have the property of being phenomenally green are as well. Smart’s adverbialism, then, though he does not appear to recognise this, is at root consistent with our propositions (1) to (3).

However there are tougher adverbial nuts available for cracking. Chisholm wrote explicitly that properties like green and centaurian (i.e. looking like a centaur) cannot be attributed to what he calls “looks.” “The relation of ‘looks green’ to ‘looks’ is that of ‘green’ to ‘color,’ not that of ‘green’ to ‘apple’” (1957, 116): that is, looks green, Chisholm argues, is not a property of looks, but is a more specific type of look. “Whatever else appearances may be … they are not ‘objects to a subject’” (117).

How far are statements like these in opposition to the position outlined in the previous sections? This hangs partly upon how we interpret “looks green.” If Chisholm just meant that experiences are not green in the same way that apples are—that they, so to speak, look green but actually are not (objectively green)—then common ground is available; likewise, we can agree with Chisholm that phenomenal properties are not properties of experiences (since, recall, experiences are not qualifiers—they are not the right kind of entity to be a property-holder in this way). On the other hand if Chisholm really meant, as it must be admitted he (unlike Ducasse) sometimes appears to, that the phenomenal quality of greenness is not a property at all—that “looks green” can in no way be construed as a property of anything—then he is opposition to premise vi) of the argument from perceiver relativity (and hence its conclusions).

The case is similar with Chisholm’s statement to the effect that appearances are not “objects to a subject” (though here the burden of interpretation is probably on the other foot). The claim might be taken to mean that experiences are not ‘objects’—that is, I suppose, do not exist as individuals at all; this would contradict proposition (1) above. Contrariwise, the claim might more reasonably be taken to mean that experiences are not objects distinct from and observed by the perceiver—they are not intermediate ghost qualifiers which are the proximate per-
ceived objects, but instead are part of the process of the perception of external material objects. With this latter claim we have no conflict.

It is possible then to adopt a form of adverbialism—call it “strong adverbialism”—which contradicts the propositions defended above (in particular, through the denial of premise vi) of my argument from perceiver relativity), and it is likely that some of the leading adverbialists committed themselves—whether deliberately or not—to such a form. Does the adverbial stance, then, provide any additional reason for denying the truth of vi), in addition to those considerations we have already rejected? It does not: on the contrary, if anything it renders its advocates even less able to deny the premise. This is because of the problem, already alluded to, that strong adverbialists face in saying in what respect different kinds of sensing are more or less similar, if not in terms of their phenomenal quality. What is it that differentiates sensing redly from sensing greenly, and that makes the former, say, a “warmer” or less “peaceful” kind of sensation, even when the objective colour is not present, if it is not the phenomenal nature of the sensation?

Frank Jackson (1976, 388; 1977, 63 ff.) built upon this problem and offered what he calls the Many Property objection to the adverbial way of talking. He begins by pointing out that visual images (for example) are not just red, but may also be, say, square and moving from left to right. Now, “sensing F-ly” is all very well, but what if there is more than one attribute? Suppose we analyse it as “sensing F-ly and G-ly.” This analysis allows the appropriate entailment from sensing a red square to sensing something red, but it fails to distinguish between “I have an F and a G image” and “I have an F, G image”—thus the adverbial analysis of the following two distinct state of affairs will be identical: sensing a red, square image at the same time as a green round one; and having a green, square image at the same time as a red, round one. The analysis in both cases is that I am sensing red-ly and round-ly and square-ly and green-ly.

One possible response to this problem would be to categorise sensings according to their positions in the visual field: the red square is, say, to the left of the green circle. But, as Jackson points out, this can’t help. Merely adding adverbs of position to the conjunction would be silly, not least because “I sense red-ly and to the left-ly and green-ly” is logically equivalent to “I
sense green-ly and to the left-ly and red-ly,” which would be the analysis of seeing green to the left of red. Alternatively, adding some specification of the location of the sensing in the visual field (say \((\exists x)[x \text{ is a part of my visual field and I sense red-ly]})\) is apparently unavailable to the adverbial theorist since it would commit her to a species of mental object: namely, parts of visual fields.

What about “I sense \(F\)-ly \(G\)-ly,” as in “he wrote astonishingly slowly”? This, unlike “he saw a red, square image,” does not entail “he wrote astonishingly” or “he wrote slowly astonishingly”; yet clearly, having a red, square image is a special case of having a red image.\(^4\)

No other, more promising, avenues appear to be open to the adverbialist: the upshot, then, seems to be that far from commonsensically clarifying matters, the adverbial framework introduces irremediable difficulties into the identification of phenomenal states, and this can hardly recommend it as a vantage point from which to attack premise vi).

Even worse, Jackson targets adverbialism for another objection—the so-called Complement objection (1976, 387–390). Consider “I have a school-age child.” Does “school-age” here qualify the child or the having of children? We can show it must be the former by this argument: suppose “I have a school-age child” and “I have a non-school-age child” are both true. Then, since nothing can be both \(F\) and non-\(F\), the predicate “school-age” cannot apply to my having of the children, and so must apply to the children themselves. Now, exactly the same structure applies to the case of perceptual experience. One can have an image which is \(F\) and one which is non-\(F\) at the same time; yet one cannot sense both \(F\)-ly and non-\(F\)-ly simultaneously (just as one cannot sing both well and badly, or run both fast and slowly). Thus, \(F\)-ness must apply to experiences and not to the having of those experiences.

The obvious reply here is to say that one can \(V F\)-ly with respect to \(A\) and non-\(F\)-ly with re-

\(^4\)Similar problems exist for Wilfrid Sellars’ talk of “sensing red-triangular-ly” (1967a), if by this he means to introduce a quite new mode of sensing, rather than a compound one: sensing red-triangular-ly could not be a special case of sensing red-ly since sensing red-ly is not even a component of that state on this view (Jackson 1976, 389). Also, why just “red-triangular-ly”? Why not “red–triangular–fuzzy–to-the-left-of-the-blue–after-image–ly”? This process could continue almost without end.
spect to B: thus one can listen happily to the strings and unhappily to the piano during a concerto. But then with respect to what am I sensing redly, if not a mental object? (390)

From all of this I conclude that the most common contemporary school of objection to sense-datum theory, the adverbial stance, fails to impact against the theses I am defending here. Under its arguably most defensible construal, adverbialism turns out to be merely a confusing restatement of my non-identity claims; in its strong form, adverbialism conflicts with the argument from perceptual relativity, which has already been shown to be sound, and adds nothing to any attempted critique of that argument’s premises.

5. CONCLUSION: AGAINST ‘FIGMENT’?

I take it, then, that this chapter—in combination with the plausible but not definitive arguments from the causal nature of perception in the previous chapter—has so far demonstrated the truth of the following three claims:

1. Some external perceived objects are non-identical with experiences, and *vice versa*, for example, strawberries are not identical with visual experiences of strawberries.

2. Some qualia are non-identical with properties of external perceived objects, and *vice versa*, for example, the actual redness of a strawberry is not identical with the quale

41 Another objection sometimes mooted is the following:

a) If it is correct to use the adverbial sensing terminology to describe the phenomenological content of sensory experience, then if S senses x-ly, he has a sensory experience of x-ness.

b) If it is correct to use the adverbial sensing terminology, then S senses x-ly just in case S senses in an x manner.

c) It is false that if S senses in an x manner (for example in a dazed manner), then S has a sensory experience of x-ness (for example, dazedness).

d) Therefore it is not correct to use the adverbial sensing terminology to describe the phenomenological content of sensory experience.

But b) is not right. “Redly” is *not* an adverb of manner, like “slowly.” Rather it describes a kind of sensing: it is more like “waltz-dancing,” which describes a kind of dance rather than a manner in which dancing is done. Perhaps, then, we should say: “to x-sense,” rather than “to sense x-ly.” But *neither* are equivalent to “in an x manner.”
of redness experienced by that strawberry’s perceiver (or, certainly, the phenomenal greyness experienced by a colour-blind perceiver).

(3) Some external perceived objects are non-identical with qualifiers, and vice versa; for example, strawberries are distinct from the individuals which have among their properties the qualia associated with perceptions of strawberries.

So, at least sometimes, phenomenal properties inhere in objects other than the external objects of perception of which they appear to be properties. The question we must ask now is: What objects do they belong to then? What actually are these mysterious qualifiers I keep talking about? Unless we want to return to substance dualism, the answer seems clear. The only other plausible physical property-holder available, once perceived objects external to the human sensory system are ruled out, are states of that system—that is to say, most plausibly, states of the central nervous system. It is hardly worth taking seriously the notion that phenomenal colours, tastes and pains might be properties of some physical object other than external objects of perception or brains—such as afferent nerve endings, for example, or volumes of thin air, or unperceived objects. It is true that the ‘substance physicalist’ can recognise properties of individuals which are not physically constituted, such as numbers, centres of gravity, or inexistent intentional objects, but again these just do not seem to be the sort of objects which can plausibly be said to be phenomenally coloured, smelly or painful.

My claim that qualia must be properties of brain states, then, relies upon the notion that the only plausible property-holding entities around are external objects and brains, and that the impulse to hypothesise extra individuals to form a basis for qualia—such as, on certain interpretations, “images” or “sense-data”—is one to be resisted. For example, one might loosely say that qualia are “properties of consciousness” (or images, etc.), but I don’t think one should

---

42 By “CNS state” I mean the fundamental, intrinsic properties of a spatio-temporal region of the CNS, that form the basis of its other properties. However, we should think of CNS states as only determining all their physical properties, as otherwise the supervenience of qualia follows from this argument (which is nowhere near strong enough to legitimately establish such a conclusion). From now on, for added vividness, instead of talking cumbersomely about “CNS states” I shall usually talk simply of “brain states.”
mean by this that consciousness is a *thing* which has properties; images do not “have” phenomenal greenness in the same way that apples have the property of greenness. Rather phenomenal consciousness is a process *made up of* qualia, and just as apples are green, *brain states are phenomenally* (though *obviously* not actually—see below) green. That is, they are (something like) sensations of greenness, in some full-blooded way—they have a property that is exactly *that token colour experience* you have when looking at an apple.

The language can be very difficult here—it’s extremely hard to avoid possible misinterpretation. For example, why do I say images aren’t experientially green (like apples are green)? Because images do not *exist* in the same way apples do; there are apples, there are causal perceptual chains, and there are brains—nothing else; no additional ‘image-entities.’ On the other hand, of course, perhaps images (if we want to say they exist at all in some way more substantial than experiences (in my standard qualia-bundle sense)) are *parts* of the brain, over a certain time-interval. Certainly, in *that* case, presumably, qualia are properties of the states of the brain that are identical with the image of the apple, and so they might be called properties of the image. So, are qualia properties of images? It depends on what you mean. Are qualia properties of the brain? Surely, on any plausible account, they are.

When someone perceives a wet dog, what that person experiences is shape qualia, colour qualia, smell qualia, etc., combined in a certain way. Once all these phenomenal properties have been enumerated, there is nothing else: there is not, *in addition*, some wet dog image—certainly not one in which the qualia inhere (subject to the proviso above). To take away the qualia tokens *just is* to take away the phenomenal consciousness.

Perhaps, though, one might still want to insist that it is nevertheless *not* the *brain* in which qualia inhere. For example, one might object, when I hallucinate a black dog it is not the brain which is black but *the (hallucinatory) dog*. Generally, it does not seem to be true that predicates used in articulating the contents of phenomenal thought are predicates which are true of or instantiated by states of the brain—rather, they are true of the entities thought *about*. Further, the dog I hallucinate, one might say, is not itself a property-token: rather, it has the ontological status of something that *has* properties (in this case, blackness). To put it another way, to de-
scribe the phenomenal features of the hallucination one typically uses nouns (such as “dog”), and nouns usually denote *individuals* (rather than properties), hence potential property-holders.

This may well seem a plausible and perhaps even familiar response, but I think any air of plausibility is misleading—or perhaps better, misdirected: the points made are plausibly true of descriptions of the *content* of hallucinations, but cannot possibly be true of the qualia themselves. Suppose you hallucinate a black dog. There is no dog and nothing relevant that is black—we want to avoid any theory (like a simple-minded sense-datum theory) which hypothesises something black or some kind of intermediate dog-substitute. That is, in short, there is nothing around that has the property of blackness. On the other hand, there are certainly a bunch of quale-tokens being instantiated, including phenomenal blackness: there is nothing black, but there *is* the appearance of blackness. Of what are these property tokens properties? Well they, also, cannot be properties of the hallucinatory dog, again because the dog is *hallucinatory*—it does not exist. They therefore must be properties of something else … and here we are back where we started. Qualia, in sum, cannot be properties of hallucinatory dogs because hallucinatory dogs do not exist.

On the other hand, the *content* of my hallucination is, let us suppose, that there is a black dog. Furthermore, by “dog” presumably we mean what we always mean and, in particular, we mean something which has properties. It is to this (non-existent) dog that I (falsely) attribute the property of actual blackness: the predicate “is black” in my account of the *content* of the hallucination, picks out, not a property of my brain, but a putative property of dogs. It may be at least roughly true to say that “the content of my hallucination is a dog,” and that “the colour of that dog is black.” But it does not follow that the *content* of my hallucination is black (phenomenally or otherwise); that *would*, I think, be a category mistake—contents, unlike dogs, are just not the kind of thing that can be coloured. Instead we must insist that properly speaking the (falsidical) content of my hallucination is *that* some dog is black.\(^4^3\)

\(^{43}\) What about the following objection? “When I hallucinate a black dog, the blackness I hallucinate is a property of the dog I hallucinate; therefore if I have ‘sense data’ when I hallucinate then my ‘sense data’ ‘include’ both properties (qualia) *and* objects, such as dogs. Now, on your story, the dogs I hallucinate do not seem to be qualia
The ultimate conclusion of the argument from perceiver relativity, then, is that qualia—by which we mean phenomenal properties—are (at least sometimes) properties of brain states (since they are properties of something, and brain states are the only plausible candidates). Perhaps this conclusion seems just uninteresting, after all that. From some standpoints, one wants to say that of course qualia are properties of the brain! Why should that need proving? Yet it is, if not a surprising conclusion, a metaphysically significant one—and, as it happens, it is a position that seems to be routinely denied, in one way or another, by many if not most of the major theorists in the field, such as Daniel Dennett, Fred Dretske, Bill Lycan, Michael Tye, Gilbert Harman, David Rosenthal, Austen Clark, Bob Kirk, David Armstrong, the Churchlands (probably), and so on. The key, of course, is what is meant by the word “qualia” in the

(since they are not properties) or qualifiers (since hallucinatory dogs are not states of the CNS); in fact, they have simply been left out of your story, and—as the language of sense data suggests—it is here that we must look for the individuals of which our qualia are properties.” My response is as follows: hallucinatory dogs have indeed been left out of my ontology since, apart from their status as inexistent intentional objects, they simply do not really exist—there are, I assume, no hallucinatory dogs populating the real world. On the other hand, I do recognise the existence of hallucinations of dogs, even qua products rather than processes. These hallucinations, however, I do not consider to be extra property-holding individuals additional to all the physical states of the CNS involved with sensation: they are physical individuals, since I rule out—more or less by fiat—non-physical individuals (a.k.a. ghost qualifiers), and the only relevant physical individuals are brain states. Thus, though hallucinatory dogs are not states of the CNS, hallucinations of dogs (at least in one sense) are. (I say “in one sense” since one can also think of hallucinations as falsidical experiences of a certain—phenomenally black and doggy—sort. This is the sort of sense in which one might want to say (loosely) that “hallucinatory dogs exist.” But by this—at least on my account—one would not mean to be reifying any doggy, property-holding individual, any more than the roughly analogous claim that “painted centaurs exist” commits one to the existence of centaurs, or even centaur-like creatures … never mind centaur-like entities which are serious candidates for having the property of whiteness. Indeed, this is just the sort of reductio from which my terminology is designed to rescue ‘sense-datum theory.’

44 “The [predicative] materialist line rejects any presentation of phenomenal colour. Here the perception is just a ‘physically acceptable’ state of the brain to which first-person concepts are applied in introspection. Nothing is red—objectively, phenomenally, or otherwise. The brain isn’t red, and the introspective concepts neither are, nor need otherwise introduce, phenomenal red. Just as tokens of the English predicate ‘red’ can represent objects as red without themselves being, or otherwise needing to present, red, tokens of the mentalese predicate ‘RED’ can represent red-perceptions as red-perceptions without themselves being, or otherwise needing to represent, phe-
claim that qualia are properties of brains—and what we mean here, the core meaning we have held constant throughout, is that qualia are the properties that (we might metaphorically say) fill our sensory fields, as considered from the first person perspective. Our conclusion, then, is that these very properties—phenomenal colours, tastes, smells, tickles and so on—turn out to be non-identical with the properties they often purport to ‘represent,’ are not themselves actually properties of the objects of perception, but are instead properties of the brain. The ‘redness’ we experience when we look at a strawberry—\textit{that} property—is what is now being attributed to brain states. Not merely the capacity to discriminate redness from greenness, not just some representation that redness is the case, or simply a state with the content that some intentional object is red—\textit{phenomenal redness}.

To make this conclusion more concrete, then, here is the ontology of the perception of a wet dog. First, there exists an everyday, physical, drenched canine with the real properties of being hairy, brown, smelly, and so on. I have nothing to say here about the metaphysics of dogs; choose your preferred account of individuals and their properties. Second, there is what we might loosely call a causal chain from the dog to the central nervous system of the perceiver. Third there is some sequence of states of the central nervous system—possibly highly dispersed and chronologically asynchronous—which make up the relevant perceptual processing. Fourth, there are the various simple and complex \textit{properties} of these states, including their location, size, electrical charge, mass, input-output characteristics, and so on. And among these properties, we are now forced to admit, are \textit{phenomenal} properties—qualia. That is, certain of these brain states have the property of being phenomenally brown, phenomenally wet-feeling, phenomenally smelly, and so on, and any self-respecting theory of mentality needs to treat these phenomenal properties as a central part of the domain it sets out to explain. This is the central positive thesis of this chapter. Fifth and finally, we are also permitted to talk about the furniture of our phenomenal mental life, including the visual image of the dog, as long as we...
steer clear of the trap of postulating extra individuals in the ontology of perception.  

Perhaps contrary to first appearances, however, this conclusion has little impact upon the solution to the problem of mind-brain relation. In particular, qualia might still be irreducibly mental properties of brain states, and so property dualism or some form of double-aspect theory might still be true; on the other hand, it is important that this result in no way proves the falsity of physicalism either. In defining qualia as phenomenal, first-person properties, you will recall, we did not assume that qualia are only accessible subjectively, from the first-person. It could still be—for all that we have said so far—that phenomenal properties are identical with, or reducible to in some as yet unforeseen way, complexes of standardly ‘physical’ properties of the brain like chemical composition or electromagnetic oscillation. Alternatively, they could be identical with some set of third-person observable properties that have yet to be discovered such as perhaps, along the lines of the theory of Penrose and Hameroff (e.g. Hameroff and Penrose 1996), the microtubular collapse of quantum gravitational ‘bubbles.’ In the final third of this dissertation I shall examine what can be said a priori about these metaphysical questions.

Having decided that qualia are properties of brain states, there is one more important thing

\[45\] I have been speaking of experiences as, roughly, sections of the stream of consciousness—as ‘bundles’ of qualia—in a way analogous, I take it, to the sense-datum theorists’ talk of “coloured patches.” On the other hand, it is possible (though I have not done so) to use terms like “images” or even “sense-data” to pick out qualifiers; for example, one might label the diffuse CNS state which has the bundle of phenomenal properties that make up the experience of the dog the “sensory image dog.” Note, however, that this is not a usage I have adopted; typically, when I speak of “images” I mean experiences, not states of the CNS.

\[46\] David Armstrong honestly professed his anxiety in 1968 that, in admitting qualia to the brain, “we are back to that bifurcation of mental and physical reality which it is the object of a physicalist doctrine of man to overcome. … To accept the view that the secondary qualities are irreducible qualia of mental items would be to abandon the whole programme of this work” (1968, 272). But we have seen that this worry, at least with respect to the argument from perceiver relativity presented here, is without basis. True, making brain states phenomenal is still consistent with property dualism—qualia do not entail physicalism … but who would have ever thought they do or should? On the other hand, we have seen that qualia (as defined here) do not necessarily entail the falsity of physicalism. Notice that Armstrong’s objection is to irreducible qualia that are properties of mental items: the qualia whose existence we have defended here are neither necessarily irreducible, nor are they properties of particularly ‘mental’ entities—they are properties of brain states, as Armstrong himself would wish.
we can say about qualia straight away. Obviously states of the brain are not actually forest green or excruciatingly loud—we simply cannot think of phenomenal properties on the model of the ‘objective’ properties they, so to speak, ‘represent,’ and to do so is to under-estimate the thesis.\(^{47}\) Instead, we must think of phenomenal properties as being such that to token phenomenal property \(F\) is to be the phenomenal sensation of \(F\). Thus, phenomenally green brain states are not green—they have the property of being the sensation of green. To be phenomenally green is not to look green but to feel green. What precisely do we mean by this? Well, we must mean something by it, since it must be the case, but exactly what is deeply puzzling. This dissertation is far from a solution to “the problem of consciousness” and, though this notion of a phenomenal property is (in my view) an immensely useful and powerful one, as I shall try to show in some of the following chapters, it is here that one can look for the still-resistant kernel of the mystery of consciousness.

That qualia are, it turns out, felt properties also has epistemological implications. Phenomenal redness is a property of the brain that feels a certain way—it is conscious in a certain manner. It may perhaps not be necessary that for qualia to exist is for them to be experienced, but that is certainly the normal case. On the other hand, as we have seen, we are under no a priori constraint to identify perceptual, intentional, content with qualia: I might conceivably be in a state indicating that some object is red yet experience phenomenal greyness (because I am colour blind), or no qualia at all (because the perception was subliminal). Further, beliefs about our own qualia are presumably to be distinguished from the mere having of those qualia—at least under all the standard (propositional, sentential or dispositional) metaphysics of belief. I might, at least logically possibly, token a phenomenal property in my brain but fail to stand in any mental relation to a proposition describing that quale, or fail to token a sentence in the Language of Thought about that quale, or fail to have any disposition to assent to claims that I just did—or

\(^{47}\) As Dennett puts it, satirically: “It may look as if the color is out there, but it isn’t. It’s in here—in the ‘eye and brain of the beholder.’ … But now, if there is no inner figment that could be colored in some special, subjective, in-the-mind, phenomenal sense, colors seem to disappear altogether! Something has to be the colors we know and love, the colors we mix and match. Where oh where can they be?” (1991, 370–371).
even am now—experiencing that quale; and so on. Thus, different sorts of ‘knowledge claim’
in this domain may come apart: to pick an extreme example, I might simultaneously perceive
that something is red, experience a green quale-token, and have no belief at all about my own
qualia. In the next three chapters I shall explore some of the implications of this view for the
traditional epistemological ‘issues’ for qualia: their putative, and peculiar, immediacy, certainty
and privacy.
APPENDIX: PROOF THAT QUALIA ARE INTERNAL PROPERTIES

Universe Of Discourse: phenomenal (but possibly also objective) properties; times; and individuals.

\(a\): a randomly chosen external perceptual object.

\(Px\): \(x\) is a perceiver.

\(S\Phi_{xyt}\): \(x\) senses \(y\) as having phenomenal property \(\Phi\) at time \(t\).

\(Q\Phi_{xt}\): object \(x\) has phenomenal property \(\Phi\) at time \(t\).

1. \((\exists x)(\exists y)(\exists \Phi)(\exists \Psi)(\exists t)((P_{x} & P_{y}) & x \neq y) & ((S\Phi_{xbt} & S\Psi_{yat}) & \sim (Q\Phi_{at} & Q\Psi_{at}))\) Ass.

There is some external object (call it ‘\(a\)’) that is sensed by one perceiver at time \(t\) to have phenomenal property \(\Phi\) and by another perceiver at the same time to have phenomenal property \(\Psi\), but that object does not have both properties at the same time.

2. \((\forall \Phi)(\forall \Psi)(\forall t)((Q\Phi_{xt} \lor Q\Psi_{xt}) \lor \sim (Q\Phi_{at} & Q\Psi_{at}))\) Ass.

‘\(a\)’ is either \(\Phi\), or \(\Psi\), or neither.

3. \((\forall x)(\forall y)(\forall t)((Q\Phi_{xt} \lor \sim Q\Phi_{yt}) \supset x \neq y)\) Ass.

If some \(x\) has the property \(\Phi\) at time \(t\), while some \(y\) does not have that property at \(t\), then \(x\) and \(y\) are non-identical.

4. \((\forall x)(\forall y)(\forall t)(\exists z)(S\Phi_{xyt} \supset Q\Phi_{zt})\) Ass.

If someone senses some \(y\) as having phenomenal property \(\Phi\), then there is something (not necessarily \(y\)) that has the phenomenal property \(\Phi\).

5. \((\exists y)(\exists \Phi)(\exists \Psi)(\exists t)((P_{b} & P_{y}) & b \neq y) & ((S\Phi_{bat} & S\Psi_{yat}) & \sim (Q\Phi_{at} & Q\Psi_{at}))\) 1 \(\exists E\)

6. \((\exists \Phi)(\exists \Psi)(\exists t)((P_{b} & P_{c}) & b \neq c) & ((S\Phi_{bat} & S\Psi_{cat}) & \sim (Q\Phi_{at} & Q\Psi_{at}))\) 5 \(\exists E\)

7. \((\exists \Psi)(\exists t)((P_{b} & P_{c}) & b \neq c) & ((S\Phi_{bat} & S\Psi_{cat}) & \sim (Q\Phi_{at} & Q\Psi_{at}))\) 6 \(\exists E\)

8. \((P_{b} & P_{c}) & b \neq c) & ((S\Phi_{bat} & S\Psi_{cat}) & \sim (Q\Phi_{at} & Q\Psi_{at}))\) 7 \(\exists E\)

9. \((P_{b} & P_{c}) & b \neq c) & ((S\Phi_{bat} & S\Psi_{cat}) & \sim (Q\Phi_{at} & Q\Psi_{at}))\) 8 \(\exists E\)

10. \((S\Phi_{bat} & S\Psi_{cat}) & \sim (Q\Phi_{at} & Q\Psi_{at}))\) 9 \&E

\(a\) is sensed at time \(t\), as being \(A\) by perceiver \(b\) and \(B\) by perceiver \(c\), but it is not both \(A\) and \(B\) at the same time.

11. \((\forall \Psi)(\forall t)((Q\Phi_{at} \lor Q\Psi_{at}) \lor \sim (Q\Phi_{at} & Q\Psi_{at}))\) 2 \(\forall E\)

12. \((\forall t)((Q\Phi_{at} \lor Q\Psi_{at}) \lor \sim (Q\Phi_{at} & Q\Psi_{at}))\) 11 \(\forall E\)

13. \((Q\Phi_{at} \lor Q\Psi_{at}) \lor \sim (Q\Phi_{at} & Q\Psi_{at}))\) 12 \(\forall E\)

14. \((S\Phi_{bat} & S\Psi_{cat})\) 10 \&E

15. \((Q\Phi_{at} & Q\Psi_{at})\) 10 \&E

16. \((Q\Phi_{at} & Q\Psi_{at})\) 15 \(\text{DeM}\)

17. \((Q\Phi_{at})\) ACP

18. \((Q\Phi_{at})\) 14 \&E

19. \((Q\Phi_{at})\) 16, 17 DS
Chapter 3: “Qualia and the Argument from Perceiver Relativity.”

20. $S_{\text{Bat}} \land \neg Q_{\text{Bat}}$

21. $Q_{\text{Aat}} \supset (S_{\text{Bat}} \land \neg Q_{\text{Bat}})$

22. $Q_{\text{Bat}}$

23. $S_{\text{Abat}}$

24. $\neg Q_{\text{Aat}}$

25. $S_{\text{Abat}} \land \neg Q_{\text{Aat}}$

26. $Q_{\text{Bat}} \supset (S_{\text{Abat}} \land \neg Q_{\text{Aat}})$

27. $\neg Q_{\text{Aat}} \land \neg Q_{\text{Bat}}$

28. $S_{\text{Bat}}$

29. $\neg Q_{\text{Bat}}$

30. $S_{\text{Bat}} \land \neg Q_{\text{Bat}}$

31. $(S_{\text{Bat}} \land \neg Q_{\text{Bat}}) \lor (S_{\text{Abat}} \land \neg Q_{\text{Aat}})$

32. $(\neg Q_{\text{Aat}} \land \neg Q_{\text{Bat}}) \supset ((S_{\text{Bat}} \land \neg Q_{\text{Bat}}) \lor (S_{\text{Abat}} \land \neg Q_{\text{Aat}}))$

33. $Q_{\text{Aat}} \lor Q_{\text{Bat}}$

34. $(S_{\text{Bat}} \land \neg Q_{\text{Bat}}) \lor (S_{\text{Abat}} \land \neg Q_{\text{Aat}})$

35. $(Q_{\text{Aat}} \lor Q_{\text{Bat}}) \supset ((S_{\text{Bat}} \land \neg Q_{\text{Bat}}) \lor (S_{\text{Abat}} \land \neg Q_{\text{Aat}}))$

36. $(S_{\text{Bat}} \land \neg Q_{\text{Bat}}) \lor (S_{\text{Abat}} \land \neg Q_{\text{Aat}}) \lor ((S_{\text{Bat}} \land \neg Q_{\text{Bat}}) \lor (S_{\text{Abat}} \land \neg Q_{\text{Aat}}))$

37. $(S_{\text{Bat}} \land \neg Q_{\text{Bat}}) \lor (S_{\text{Abat}} \land \neg Q_{\text{Aat}})$

38. $(S_{\text{Bat}} \land \neg Q_{\text{Bat}})$

39. $(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}})$

40. $(\exists \eta)(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}})$

41. $(\exists \lambda)(\exists \eta)(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}})$

42. $(S_{\text{Bat}} \land \neg Q_{\text{Bat}}) \supset (\exists \lambda)(\exists \eta)(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}})$

43. $(S_{\text{Abat}} \land \neg Q_{\text{Aat}})$

44. $(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}})$

45. $(\exists \eta)(\exists \phi)(S_{\phi_{\text{bat}}} \land \neg Q_{\phi_{\text{at}}})$

46. $(\exists \lambda)(\exists \eta)(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}})$

47. $(S_{\text{Abat}} \land \neg Q_{\text{Aat}}) \supset (\exists \lambda)(\exists \eta)(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}})$

48. $(\exists \lambda)(\exists \eta)(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}}) \lor ((\exists \lambda)(\exists \eta)(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}}))$

49. $(\exists \lambda)(\exists \eta)(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}})$

$a$ is sensed as having some phenomenal property that it does not in fact possess.

50. $(\exists \eta)(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}})$

51. $(\exists \phi)(S_{\phi_{\text{cat}}} \land \neg Q_{\phi_{\text{at}}})$

52. $S_{\text{Acat}} \land \neg Q_{\text{Aat}}$

53. $\neg Q_{\text{Aat}}$

54. $S_{\text{Acat}}$

55. $(\forall \lambda)(\forall \eta)(\exists \phi)(S_{\phi_{\text{cat}}} \supset Q_{\phi_{\text{at}}})$

56. $(\forall \eta)(\exists \phi)(S_{\phi_{\text{cat}}} \supset Q_{\phi_{\text{at}}})$
Chapter 3: “Qualia and the Argument from Perceiver Relativity.”

57. \((\forall t)(\exists z)(S_{Acat} \supset Q_{Azt})\)

58. \((\exists z)(S_{Acat} \supset Q_{Azt})\)

59. \(S_{Acat} \supset Q_{Abt}\)

60. \(Q_{Abt}\)

61. \((\forall y)(\forall \Phi)(\forall t)((Q_{\Phi bt} & \neg Q_{\Phi at}) \supset b \neq a)\)

62. \((\forall \Phi)(\forall t)((Q_{\Phi bt} & \neg Q_{\Phi at}) \supset b \neq a)\)

63. \((\forall t)((Q_{Abt} & \neg Q_{Aat}) \supset b \neq a)\)

64. \(((Q_{Abt} & \neg Q_{Aat}) \supset b \neq a)\)

65. \(Q_{Abt} \supset \neg Q_{Aat}\)

66. \(b \neq a\)

67. \(Q_{Abt} \& S_{Acat}\)

68. \((Q_{Abt} \& S_{Acat}) \& b \neq a\)

69. \((\exists t)((Q_{Abt} \& S_{Acat}) \& b \neq a)\)

70. \((\exists \Phi)(\exists t)((Q_{\Phi bt} \& S_{\Phi yat}) \& b \neq a)\)

71. \((\exists \Phi)(\exists t)((Q_{\Phi bt} \& S_{\Phi yat}) \& b \neq a)\)

72. \((\exists x)(\exists y)(\exists \Phi)(\exists t)((Q_{\Phi xt} \& S_{\Phi yat}) \& x \neq a)\)

There is something which has some phenomenal property \(\Phi\), a phenomenal property that \(a\) is sensed as having, and this thing is non-identical with \(a\).
Chapter 4: The Immediacy of Qualia Apprehension

Qualia (and experiences) have typically been thought to stand in a very different epistemological relation to the knower than the external furniture of the world. The “raw feels” of thought are often said to be “given,” while what we might call the content of that thought—usually, claims about the external world—is only more or less doubtfully true; and this is (or at least was) usually said to be because we are “directly” or “non-inferentially” confronted by qualia or experiences, whereas all other properties or objects are only mediately ‘connected’ to the perceiver. In the past this was often thought to be a theoretical virtue of the notion of ‘sense data’ or ‘qualia’—something that blocked absolute scepticism, and offered the (perhaps ultimately illusory) hope of building a body of certain knowledge about the world from the fundamental building block of the quale. In today’s context, however, this epistemological asymmetry between qualia and, roughly, the rest of the universe is a serious barrier to acceptance of the notion: the putative “givenness” of qualia is often thought to place them in the realm of the irremediably non-physical, subjective, or generally metaphysically weird. Indeed, since Givenness is thought to have recently been proved a Myth (see Chapter Six, section six), qualia and experiences—qua uniquely “given” internal particulars—have also been called into question.

What is this “givenness” supposed to amount to? Just what is special about the perceptual status of qualia? There are, I think, three basic notions involved here: the immediacy of our apprehension of qualia (or, less accurately, our “direct perception” of qualia); the certainty of our apprehension of qualia; and the privacy of our apprehension of qualia. For example, H.H. Price, whose Perception is perhaps the most highly developed presentation of the sense-datum theory, held that “acquaintance with sense-data” is an indubitable apprehension and (or because) it is “an immediate ‘confrontation’ of the mind and sense-data” (1950, 149), and he went on to ask rhetorically, “Is not each sense-datum private to the mind that senses it?” (274).

1 The contrast between “raw feel” and “content” is actually a rather complex one, as may appear in due course. Here I mean primarily to contrast the doubtfulness of the truth of what is represented by sensation with the “givenness” of the experience itself.
Moore, too, asserted in his seminal lectures of 1910 that sense-data are “directly apprehended”: “...[W]hen I see this whitish colour, I am directly apprehending this whitish colour: my seeing of it, as a mental act, an act of consciousness, just consists in my direct apprehension of it” (1953, 33).

These three quasi-epistemological traits have historically been treated as very closely interconnected: for example, ‘sense-data’ were often thought to be certainly known because they are immediately or directly known, and further thought to be private in virtue of being only directly known. Likewise, that sense-data are known with certainty has been considered evidence that they are indeed known immediately (as is, in a sense, their privacy—thier perceiver-relativity). In this chapter and the following two I shall try to show that all of these epistemological facets—when properly, and not too ambitiously, understood—arise naturally from the metaphysical status of qualia as being phenomenal properties, where this is understood in just the way we have established in the previous chapter. Further, I hope I shall demonstrate that these epistemological features are nothing to be afraid of. Very frequently, attacks on the existence of qualia—from Hirst’s chapter on “The Perplexing Nature of Sense-Data” (1959, Ch. 3) to Dennett’s “Quining Qualia” (1988)—have depended upon formulating some purportedly standard characterisation of qualia and then showing that “there is nothing to fill the bill” (Dennett 1988, 74); usually, immediacy, certainty and privacy are prime among the qualities that, the objection runs, no actual ‘entity’ possesses. We shall here show that this kind of objection has no basis.

It will sometimes be important, as in earlier chapters, to distinguish between sensing (the having of, say, desk-like or violin-like experiences) and perceiving (the seeing of actual, external desks or violins); all of the epistemological qualities discussed here apply only to sense and not to perception. I shall also continue to use my terms of art—“qualia,” “experience,” “qualifer,”

---

2 Really, these might be thought of more accurately as metaphysical traits that have epistemological consequences. I will use “epistemological” loosely here as just meaning something like “concerned with the study of knowledge.”

3 This does not, however, mean that there are two kinds of sensation, perceptual and sensory. Nor do I wish to
etc.—in the manner prescribed in Chapter One.

Within the domain of sensing, it is crucial to discriminate between what I shall call *phenomenal awareness* of an experience or quale, *attention* to an experience or quale, and *propositional awareness* that one is now experiencing or has experienced some experience or quale. The first has the striking characteristic, which is the result of the analysis in Chapter Three, that it is identical with the having of a quale: to experience some quale just is to be phenomenally aware of that quale—in fact, I take it, the second half of this identification is best seen as simply a reformulation of the idea expressed in the first half. This kind of awareness is wholly non-linguistic (animals can have experiences too; there is no reason to think that aphasia would rob us of our experiences), and distinct from the having of *beliefs* (or other cognitive attitudes) about qualia.

---

4 The distinctions I am making here are fairly rough and ready—for the moment I am concerned only to try and forge ahead and make progress. One example of this roughness is my claim that phenomenal awareness is non-linguistic. I will admit later that qualia are not cognitively impenetrable; thus one’s linguistic abilities might well—indeed, surely do—affect the experience of, say, listening to spoken English. This is problematic in the case of language in particular if one holds, in addition, that *understanding* the words is *part* of the experience of them—that, in some sense, the very experience itself is linguistic (rather than merely acoustic). I shall simply side-step this very difficult issue, and hope that, at any rate, my claims here hold for experiences which are not specifically experiences of *language*. 
The second kind of awareness—attention—is (in the sense I am using it here) also non-linguistic: it picks out those qualia which are being especially attended to at any one time: roughly, for example, those in the foveal area of the visual field. Thus, for instance, the whole visual field will typically, at any one time, contain both ‘clear,’ attended-to visual qualia, and ‘fuzzy,’ peripheral qualia which are experienced but not attended to. The third type of awareness, propositional awareness, involves taking a cognitive attitude (such as belief that, or surprise that) toward a quale or experience under some description—for example, believing that one is seeing a red patch.

Epistemologically speaking there is a corresponding hierarchy, from sensing to attending to believing. However, only the last of these is strictly within the domain of traditional epistemology—the type of thing, for example, that can be a justified true belief, where what is believed is identifiable with some proposition. What is apprehended in sensing and attending

---

5 This does not correspond with the standard usage of the same term in experimental psychology, though I think it is not totally dissimilar to the meaning given to the term by those who introduced it into psychology: figures like William James, Edward Titchener and Wilhelm Wundt, who seem to have thought of attention as the making more vivid or clear the imagistic, phenomenal contents of consciousness. Modern psychology, however (since Broadbent 1958), is more influenced by communication theory and computer technology, and conceives of attention as a typically subliminal information filtering process—a selective analysis of perceptual inputs by a limited-capacity information-processing system. This, for example, is the usage in Glass and Holyoak 1986, 33ff.

6 C.I. Lewis, influentially, thought that something like a phenomenal/propositional distinction is crucial here. “To suppress it altogether, would be to betray obvious and fundamental characteristics of experience. If there be no datum given to the mind, then knowledge must be contentless and arbitrary; there would be nothing which it must be true to. And if there be no interpretation or construction which the mind itself imposes, then thought is rendered superfluous, the possibility of error becomes inexplicable, and the distinction of true and false is in danger of becoming meaningless. If the significance of knowledge should lie in the data of sense alone, without interpretation, then this significance would be assured by the mere presence of such data to the mind, and every cognitive experience must be veracious” (1929, 38–39). See, however, the section of the ‘Myth of the Given’ in Chapter Six.

7 “The making of a judgement necessarily involves the application of concepts (the concept of pain, for example, or thirst), whereas the mere having of sensations and feelings does not require the application or even the possession of any concepts at all. The sensation and the judgement are therefore two distinct things, contingently related at best” (Paul Churchland 1979, 98). Dretske, similarly, argues for the “difference between a concept-free mental state (e.g., an experience) and a concept-charged mental state (e.g., a belief)” (1993, 773); his discussion on
cannot be identified with any proposition, though it may be more or less fully described by some proposition (just as, say, the governmental system of Canada is not a proposition, though it may be described by one). Nevertheless, unlike governments, sensing and attending, because they are phenomenal, may be considered a form of knowing—and I think typically (though not always) are so considered in this domain, where “acquaintance” or “immediate apprehension” are treated as epistemological concepts, and where it is routinely asked whether qualia themselves are, say, directly known. I shall use the term “apprehension” neutrally, as covering all three varieties of “knowing.”

This topic in the first section of this paper (called “Conscious Experience”) is very useful.

8 C.I. Lewis, for example, denied that non-propositional awareness can count as any kind of knowledge (1929, 37–39), as, famously, did Wilfrid Sellars (e.g. 1956). (Though Sellars and I would in fact agree that the nature of sense experience gives to conceptualisation more than simply Rylean sensory stimulus, albeit less than the content of full-blooded knowledge claims suitable for the derivation of more knowledge—see, once again, Chapter Six, section six.)

9 This taxonomy together with the general theory of qualia presented in this dissertation, it seems to me, has relevance for the so-called Knowledge Argument about qualia, which demands to know whether there are there “subjective facts” (which must therefore escape standard, objective physics) or if instead there are just different ways of knowing “ordinary” physical facts about the mind-brain. We can agree with the standard claim that “knowledge involves the mode under which the knower represents the fact known, and … this is no less true for mental facts than for ordinary physical ones. … [K]nowledge that one is having this sort of sensation differ[s] from knowledge that one is in brain state so-and-so, even though one’s having this sort of sensation just is one’s being in brain state so-and-so” (Lycan 1990, 113, 120). However “predicative materialists” (Raffman 1995) like Lycan hold that the relevant neural property causes the tokening of a first-person introspective concept—for Lycan, a mentalese predicate (such as red-experience). My own position falls closer to that of Horgan (1984a) and Papineau (1993) (and perhaps Tye 1995), insofar as they hold that the phenomenal colour property one experiences in seeing ripe tomatoes is just a physical property of one’s brain apprehended from “the first-person ostensive perspective” (Horgan 1984a, 151). The content of this perspective is expressible in a judgement of the form, e.g., “seeing ripe tomatoes has this property,” where “this property” refers to the relevant neural property. I differ from this “demonstrative material line,” however, in that for most of these theorists, we already know a priori that “the mental demonstratives provide our only access to our first-order states” (Raffman 1995, 304). For me, the phenomenal properties are first-person introspections. Mental demonstratives, or predicates, are additional introspective states; mental demonstratives only find their reference because we also ‘know’ to what they refer. (This also prevents mental demonstratives from failing to differentiate between different phenomenal states. “In its crudest form, the demonstrative line has it that we introspectively think of all our experiences as, simply, this
I will now consider to what extent, and in what ways, qualia are immediately known, certainly known, and only privately known. I shall generally proceed through analysis of the concepts involved—often, getting clear about exactly what property we are attributing to qualia allows us to judge the appropriateness of that attribution. This chapter will focus upon the immediacy of qualia. What exactly, then, is it to be “immediate” or “directly apprehended”? As many as three (more or less non-explicit) accounts of immediacy seem to me to have had influence in the debate:

1) An apprehension is “immediate” if there is no intervening entity between the sensing agent and the apprehended entity. For example, Berkeley classically defined direct (or immediate) perception as perceiving “without the intervention of others.”

2) An apprehension is “immediate” if the sensing agent adds nothing to, or has no direct control over, this apprehension. This kind of thought might be behind why qualia are labelled “given.”

3) An apprehension is “immediate” if it is not the result of a process of inference.

(property or experience). In other words, it renders all experiences introspectively identical” (Raffman 1995, 303,).

Horgan seems to recognize this point. As he puts it: “Does (4) [‘Seeing ripe tomatoes has this property’] by itself convey the information which Mary expresses (4)? I think not. Rather, since (4) employs an indexical term essentially, it seems that in order to obtain the information which Mary expresses by (4), a member of Mary’s audience would have to experience phenomenal redness himself, and would have to know that Mary is using ‘this property’ to designate the same property that he experiences. Knowledge about what qualia are like cannot be obtained by descriptive means alone, but requires the experiencing of those qualia” (Horgan 1984a, 151n). Raffman objects that “The trouble is that the physical property instantiation cannot supply the requisite content from a first-person introspective point of view” (1995, 305). This is exactly what I deny.

10 C.I. Lewis defines the given as “that element which we are aware that we do not create by thinking and cannot, in general, displace or alter” (1929,48).

11 This can be seen as a special case of 2, but I shall follow the threads of the literature and separate it out here. There was also a fourth, quite distinct, tack at one time (Moore 1905, Broad 1965) which involved arguing that, while we only ever perceive a part of any external object (such as their facing side), we can define “immediate” objects of perception as exactly those we perceive all of at any one time. I shall not consider this option here, as it is hard to cash out without the thesis of the ghost qualifier, and since it is generally considered a more than usually implausible option (see, e.g., Jackson 1977, 13 ff.).
I shall show that the first of these notions is rather confused, but interpreted charitably is true of qualia; the second unnecessarily strong and probably false of qualia; and the third—the central meaning in the ‘sense-datum’ literature—is true of phenomenal apprehension of qualia and experiences, but less certain for propositional awareness.

1. NO INTERMEDIATE ENTITY

This sort of usage typically arises from consideration of the problems of perception, and standardly contrasts our apprehension of experiences with that of external perceived objects. On this view, an individual \( a \) “directly apprehends” an object \( o \) if \( a \) apprehends \( o \) without apprehending any intermediary \( i \). By contrast, usually, apprehension is “indirect” if it is only in virtue of apprehending \( i \) that \( a \) apprehends \( o \). Thus, experiences are said to be directly apprehended, while external perceived objects are apprehended only indirectly. However, though apparently straightforward, this dichotomy conceals unfortunate assumptions that confuse and distort the issues involved.

Though I have used the more neutral term “appréhension,” the distinction is frequently understood on the model of perception; the problem with this is that paradigm cases of perception—such as seeing a table or smelling a rose—always involve at least three separate entities or events: the perceiver, the act of perception, and the object or event perceived. This trichotomy tends to suggest the following sort of mistaken picture: in the apprehension of external perceived objects four distinct entities or events are involved, and two of these might almost equally well be called “perceived objects.” There is the perceiver, the act of perception, the experience, and the external perceptual object. Hence, it was sometimes said, (1) the perceiver (2) “perceives” (3) the experience, and from this, perhaps, infers the existence of (4) the external perceptual object. And experiences are different things than perceptions—they are, instead, what is “directly” perceived. Thus Bruce Aune could speak (critically) of a standard view which sees experiences as “a kind of object standing between us and the world” (1967, 6).

It should be fairly clear that this picture is almost certainly deeply misleading, and at any rate
Chapter 4: “The Immediacy of Qualia Apprehension.”

is not entailed by the view that qualia are phenomenal properties of states of the CNS … it is not necessarily any part of what it is to be a qualophile. Consider: what is supposed to be meant by the “perceiver” in the multi-part picture? Normally, we think of persons (and usually animals) as perceivers, but this cannot plausibly be what is meant here since qualifiers (the ground of experiences) are parts of persons (since they are states of their CNS), not distinct entities from them and certainly not standing “between” them and the external world. Perhaps, then, we are to think of the perceiver as another part of a person, which lies somehow behind the experience in the chain of perception: but if this is so, there are no currently plausible candidates for such an entity. The notion of a Cartesian soul observing perceptual experiences from ‘behind’ the pineal gland has no currency today, and neither has the more prima facie plausible but empirically bankrupt notion of a ‘perceiver-module’ in the brain that somehow ‘observes’—or even brings together in consciousness—the experiences. Instead the most plausible basic model of perception, it seems to me, (at least for the non-qualophobe) is the following: persons are perceivers; they perform acts of perception; these acts of perception involve a neural information-processing process; and typically, some stages of this neural process are phenomenal—that is the act of perception involves a perceptual sensation. These perceptual sensations are what we have been calling experiences, and are made up of qualia: they are not themselves “perceived” but are phenomenally apprehended (and possibly attended to and propositionally apprehended as well). The “perception” and the experience can therefore be more or less identified; they are not two different events or objects, but the same thing (or at least, the latter is a part of the former). The external perceptual object is not a fourth thing but (in the terms of this debate) a third … and is the only “perceptual object” properly speaking. The experience is not perceived or observed, but phenomenally apprehended—to possess an experience just is to have (a certain sort of) phenomenal awareness.

---

12 See Dennett 1991 for an especially persuasive attack on the “Cartesian theatre”; Ryle (1949 and 1954) also attacks the concept of an internal observer.

13 How, then, are we to distinguish between the advocate of “direct perception” and that of an “indirect” theory? I argue elsewhere (1997) that each must either deny the truth of this basic model (such as by insisting that experi-
Experiences and qualia are indeed, then, directly apprehended in this sense. A perceiver apprehends qualia without the intervention of any “intermediary.” There is a relevant contrast with the perceived objects: qualia (and experiences) are phenomenal—that is, we saw in Chapter Three, to instance those properties is to experience them; this is not true of the objective properties of the perceived objects. Objective brownness, squareness, and noisiness are not apprehensible by possession—there is nothing it is “like to be” a stereo speaker. Thus, there are two general ways of coming to apprehend a property; either by instantiating it, if it is a phenomenal property, or by having an appropriate perceptual system causally linked to it in certain ways. Qualia are apprehended in the first way; objective properties in the second. We might, therefore, wish to call objective properties “indirectly” apprehended (by contrast with qualia), since they are mediated by a causal chain in a way that the apprehension of qualia is not. Furthermore, there may be a sense in which it is “in virtue of” the experience that one apprehends the external perceptual object, since the experience is the phenomenal part of the act of perception.

On the other hand, as we have seen, we should avoid saying that some object interposes itself between individuals and during perception; nor need we insist that both and are perceived, one “directly” and one “indirectly”: only the external perceptual object is perceived, and the experience is part of that perception. Finally, I think it would be incautious to suggest that it is “only in virtue of” that may be apprehended, since there are plentiful instances of unconscious perception that do not involve experiences (subliminal cues; proprioceptive perceptions; blind-sight; possibly, lingering perceptual states that are temporarily ‘forgotten,’ such as the tactile feeling of being seated or having a slight headache; and so on) and since it is arguably empirically possible—though I think unlikely—that experience could be epiphenomenal to

\[14\] Or even if, weirdly enough, objective squareness is phenomenal (I haven’t proven it isn’t), it is not apprehended in that way by us.

\[n.\]
perception (this issue will be discussed in Chapter Eight).

2. BEING UNINTERPRETED

This second view of directness treats it as being roughly synonymous with “raw” or “uninterpreted.” There was a lot of talk by such as Price and Russell, for example, of the apprehension of ‘sense-data’ as being “intuitive” or “without intellectual processes,” and sometimes it appears as if they meant to make the strong claim that sense-data have no dependence at all upon any cognitive processes, conscious or unconscious. Thus when Price explicitly confronts the position of the Cambridge logicians of the 1920s and 1930s, who argued that all apprehensions are apprehensions of facts—something of the form “that \( A \) is \( B \)” or “the \( B \)-ness of \( A \)” … apprehensions not of red, round patches, but that the patch is red and round—Price writes that, if that were so, it would be that “we are not passively ‘receiving’ or (as it were) swallowing; we are actively thinking—judging and classifying—and it is impossible to do less than this” (1950, 7): this he denies. Likewise, C.I. Lewis writes of experience as being “that element [of thought] which we are aware we do not create by thinking and cannot, in general, displace or alter” (1929, 48).

It is true, as Price suggests, that experiences are not themselves conceptual (at least not in any linguistic way—see below). They are, as he says, not descriptions (or any other quasi-linguistic entities) but experiences—“particular existent[s] having a perfectly specific nature (whether describable by us or not) … of just exactly this shade of colour and … just this shape and no other” (1950, 149–150). On the other hand, it is much less clear that experiences are not in any way dependent upon, or affected by, concepts, or other cognitive elements—it is not so clear that experiences are, so to speak, “passively received” from the external world. If this strong claim was ever deliberately and literally made on behalf of ‘sense-data’ and their

---

15 “To say that they are non-conceptual is, of course, not to deny that they can be referred to and characterized by use of concepts, or even, directly responded to by concepts in direct self-knowledge. ‘Non-conceptual’ does not mean non-conceptualized” (Sellars 1965, 14).
ilk, it was, quite likely, mistaken (or at least over-ambitious).

In today’s terms, the question might be expressed by asking whether the apprehension of experiences is *cognitively penetrable*—that is, whether experiences are affected by changes in the general cognitive structure of the sensing mind. For example, much of my *behaviour* is cognitively penetrable since changing the content of my cognitive scheme will affect it—my reaction on being offered a large quantity of money will depend upon, for example, whether I believe it is stolen or counterfeit, whether I am deeply offended by such actions, whether I desperately need money to pay for essential medical treatment, and so on. On the other hand, my reflex behaviour on being tapped on the knee with a rubber mallet is (it is usually said) not in any way affected by my cognitive states—it is cognitively *impenetrable*.

It may well be that the apprehension of experiences is, in fact, cognitively penetrable. Our cognitive states, such as our beliefs, desires and expectations, may perhaps affect the pre-conscious processing of, at least, perceptual information such that they in part determine the resultant sensory experience. Thus experimentally, as is well known, changes to cognitive schemes can bring about changes in reports of sensory experience. For example, typically, human skin does not appear to us to change colour appreciably when people move through different light conditions; however, subjects can be shown that ‘in fact’ human skin has a pronounced greenish tinge when people are standing in light filtered through the leaves of a tree (by being shown a photograph in which the tree is cropped out, for example). After becoming convinced of this, subjects typically report that they now (for a while) *experience* this greenness of human skin tone when perceiving people under trees.  

---

16 See Pylyshyn 1984, especially xvii. This is not the only way of phrasing this, or similar, questions; Patricia Churchland and Terrence Sejnowski, for example, speak instead of the influence of “abeyant” neural representations upon “occurrent” ones (1992, 142), and a similar distinction is sometimes intended by the contrast between “top-down” and “bottom-up” contributions to perceptual processing (as in, for example, Glass and Holyoak 1986, 125–126). The Jerry Fodor–Paul Churchland debate about the “rigidity” versus “plasticity” of perception covers similar ground (Fodor 1984, Churchland 1988).

17 “I used to think, probably wrongly, that if I was not the first person to use exactly this example then it was the famous Dr. Land, of Polaroid camera fame” (John Baker, personal communication). This phenomenon is some-
However, this kind of report is ambiguous. It could be that what has changed—what is cognitively penetrable—is the experience itself, including the skin-colour quale; or it could be that what has been influenced is only the subject’s judgements ‘about’ that experience, their conceptualisation of what is being seen, and that the experience remains unaltered. That is, we can still ask, has there been a change in both phenomenal and propositional awareness, or only in the latter?

In my view, this is an extremely complicated matter, and a genuinely difficult issue for the qualia theorist. Daniel Dennett has made great play of this kind of problem (Dennett 1988, 1991; Dennett and Kinsbourne 1992), and suggested that it forces a kind of reductio of the very notion of qualia. Though I do not think he is ever totally explicit on this point, he apparently intends to argue roughly as follows: that we cannot tell the difference between changes in qualia and changes in cognitive attitudes; that therefore there is no difference between the two kinds of change; and that therefore the notion of qualia should be dispensed with in favour of talk merely of (roughly) cognitive attitudes. If beer is an acquired taste, Dennett asks (1991, 396) then what changes as it is acquired: the taste-of-beer qualia, or one’s “evaluation function” over those qualia? There is, he urges, just no correct answer to this question, and so no proper distinction to be made between, as it were, tastes and likings; we can talk sensibly about changes in whether or not we like beer, but not about changes in some phenomenal taste of the beer—in “what it is like to drink” beer.

Though I think Dennett has tapped a rich mine of difficulties for the qualia theorist in this area, I disagree with the conclusions he draws. I will comment upon Dennett’s position on qualia in much more detail in Chapter Eleven, and for now—for the sake of focus—will just make a few points specifically relevant to the issue of the ‘uninterpretedness’ of qualia.

That introspective reports of changes in perceptual experience are insufficient to allow us to times called “memory colour” in the psychological literature—the tendency to see the colours one expects to see. This, and several other similar experimental cases of apparent cognitive penetration, are dealt with in the context of ‘sense-datum’ theory by Hardin 1988, 96–109. Other examples that can be found here are simultaneous contrast (which intensifies large colour differences and minimises small ones); the von Bezold spreading effect (which essentially does the opposite); and gestalt “flipping” effects upon shading, line length, and so on.
discriminate between alterations in phenomenal and propositional awareness (if they are), does
not entail that there is in principle no way of doing so. In particular, suppose that qualia are
‘physical,’ third-person identifiable properties of states of the CNS, and suppose cognitive
states—such as beliefs and evaluations—to be some other kind of physical state: say, distributed
representations possessed by neural networks, sentences in the Language of Thought coded
into certain neuronal arrays, or perhaps even holistic dispositions of the whole brain to ‘relax’
into particular states. In such a case, presumably it would be perfectly possible to tell whether
or not experiences had altered over time or not, even if the experiencers themselves were unable
to judge reliably on the basis of introspection, by examining in detail the relevant physical
properties of the sensory processing in question. If phenomenal bitterness is one kind of
brain property, and phenomenal tasting-of-hops another, and if those properties can be identi-
ified by third-person examination of the brain, then a change in sensory experience from one
to the other can be reliably picked out. So, in order to deny that phenomenal change can be
isolated, one must first deny that qualia are identifiable properties of brain states; on the other
hand, on the assumption that qualia are properties of neural states, it is perfectly coherent to
suppose that changes in cognitive schema might bring about changes in experiences either with
or without changes in propositional awareness.

Whether or not experiences are indeed cognitively penetrable in this way is, surely, mostly an
empirical question … and one that, perhaps, won’t be definitively answered until qualia can in
fact be scientifically identified in the brain. However there is currently some relevant empirical
evidence which seems to me to suggest that qualia are cognitively penetrable. As I have already
suggested, it is nowadays quite uncontroversial that identical sensory stimuli can give rise to
quite different perceptions. However, perhaps unfortunately, favourite examples in the litera-
ture are things like reversible Gestalt figures (such as the Necker cube or the old-man/rat) or

18 Unfortunately, though I think some scenario like this is likely, I don’t know that it is in fact the case. It might be
that the “subvening base” (this term will be more fully explained in a later chapter) of the cognitive state is iden-
tical with that of the experience (i.e. with the qualifier). However, all I need to make my point here is the (plausi-
ble) possibility that they are distinct.
the different perceptions of English speakers and German speakers in hearing the phoneme sequence \Empedocles \lēpt\ ("Empedocles leaped" and "Empedocles liebt"); these examples do little to remove the ambiguity between changes to experience and judgement. Other instances of the same phenomenon are available, however, which are more clearly phenomenal: for example, experiences of *shape* can be seen to be cognitively penetrable. The human visual system normally processes shading gradients as consistent with the assumption that things are generally lit from above; hence, “certain shaded contours, such as moon craters, could be seen either as convex or concave, but they are automatically seen one way or the other according as the lighting assumption dictates” (Churchland and Sejnowski 1992, 145). This default perception may *change* if, for example, one comes to learn that the objects are in fact craters, or if one is attentive to the fact that they are actually lit from below. That is, whether an experience is of a concave or a convex shape—which, arguably, is more clearly a phenomenal difference than just a propositional one—can be affected by certain beliefs in one’s cognitive scheme.\(^1\)

Further, it is thought that assumptions about whether objects are concave or convex are embedded in relatively *early* stages of the visual information processing and affect “the very manner in which the peripheral signals are processed in visual cortex” (Churchland and Sejnowski 1992, 146); this suggests that the convex or concave nature of an experience is ‘decided’ well *prior* to the phenomenal stages of the perceptual process—it is plausible, therefore, to suppose that phenomenal awareness will be affected … that *different* qualia will be conjured depending on whether the object is seen as convex or concave.

Finally, there is some experimental evidence to suggest similar defeasible “rules of thumb” in the visual systems of other animals, such as cats, monkeys and owls (Ramachandran 1988, 1990a, b). Insofar as such animals can be said to have cognitive systems, and insofar as these

\(^{1}\) Paul Churchland has said that, in addition to the plasticity of perceptual judgement, he is now willing to defend the plasticity of what he calls “sensation” itself (e.g. 1988b). He provides another suggestive example: take a folded monochromatic rectangle, placed such that one of the inside faces is in slight shadow relative to the other inside face; look at it with one eye closed (to defeat stereoscopic clues), and invert its orientation as if it were a Necker cube so that the middle fold appears closer than the outside edges; the result will be that the shadow effect will disappear, to be replaced with the sensation of two differently coloured surfaces.
rules of thumb (like the one that says that light sources are typically above) can be affected by elements of their cognitive scheme (since they are defeasible), then thus far the perceptual experiences of these animals can be said to be cognitively penetrable. Since these animals do not have language, and so presumably cannot be propositionally aware at all, there is no chance of a confusion here between the two kinds of change of awareness.

Perhaps, then, we are justified in suggesting that there is little reason to think qualia direct in the sense of “uninterpreted,” and some reason to suppose otherwise. Before leaving this question behind, however, I should note that, as Dennett has noticed, the issue provides an entrée onto a bundle of conceptual complexities to the notion of qualia that I am ignoring here. My simple three-way taxonomy of awareness—phenomenal, attentive and propositional—is, I hope, adequate for the purposes of discussing the immediacy and certainty of qualia, but is almost certainly far too simple as a description of perceptual awareness in all its subtlety. Yet, it seems to me, fixing up a more adequate account of awareness is a large and difficult task which should, by rights, respond to a huge body of literature—from Husserl to Marr—and which should probably be guided by sophisticated empirical findings in the neurology and psychology of vision; I hope I am justified, therefore, in deferring this task to another day. Here, let me just admit that my distinction between merely experiencing and experiencing with attention—roughly, between the focus and the peripheries of a sensual field—is one which seems manageable in this simple form but, thought about in detail, is difficult to

---

20 The emphasis here is on “propositionally.” One would not want to be quite so dogmatic about whether animals can, in some sense, “believe that”—see the next paragraph.
21 One consequence of note of this conclusion, emphasised by Paul Churchland, would be that observation judgements are not theory neutral, at least not in the traditional sense that their “truth is not contingent upon the truth of any general empirical assumptions, that is, … free of potentially problematic presuppositions” (1988b, 141).
22 C.I. Lewis’s chapter on “The Given Element in Experience” (1929, Ch. 2) includes a nice discussion of this issue as found in some of the historical literature (including Bergson, Berkeley, Plato, the post-Kantian idealists, and the American “new-realists”) and goes on to paint his own detailed and subtle picture. Armstrong 1981, Dretske 1993, Güzeldere 1995, Lycan 1990 and Rosenthal 1990 are among the more recent thinkers on this problem.
make concrete; and the distinction between experience and propositional awareness is surely too simple-minded an opposition to capture the deeper reality of things. This second problem is especially relevant to our discussion here. It might be that we are really dealing with a much greater range of elements of awareness, such as, perhaps:

1) Experienced shapes, colours, sounds, smells, and so on.

2) Some non-linguistic form of “seeing that” or conceptualisation—the kind of thing that hawks do when they recognise mice, for example, or infants when they acquire the “notion” of an object. Further, possibly this is part of the experience itself—perhaps it feels different for me to see a house than for, say, a chimpanzee (which, let us suppose, experiences the same colour patches as me) to do so.

3) Some non-linguistic attitude towards or evaluation of what is experienced: that is, is the likeableness of a taste somehow part of the experience? Would the sound of a rattlesnake be a different experience if it was a pleasant, instead of a fearful, sensation? Is the taste of Dennett’s beer actually different now that we like it, or is it the same taste but now we have adopted a different attitude towards it? How do emotions fit in here: does it feel different to hear Beethoven’s Ode to Joy movement when we are happy and when we feel violently angry?

4) Contributions to experience that are dependent upon having a language: in particular,

---

This is closely related to the “problem of the speckled hen,” discussed by, among others, Ayer 1940, 117; Hirst 1959, 52–55; Price 1941, 280ff.; and Chisholm 1942, 368ff.. In essence, if we experience something—such as a hen—at first vaguely, and then turn all our attention onto it and experience it much more clearly, we can ask was it “there” in all that detail already but, say, the number of speckles was not noticed; or does the change in attention so to speak bring about the greater clarity? That is, in the old language, “can sense data have properties they do not appear to have”? Ayer answered “no” to this question, and argued that we have two different ‘sense-data,’ one vague and one speckled; Price and Chisholm both disagreed with this, and preferred to talk in terms of persistent sense-data which are attended to in different degrees. I side with Ayer on the issue: qualia are phenomenal properties of stages of perceptual processing, and change constantly as the process of perception changes—the experience of the attended-to speckled hen is a new and more recent experience than that earlier one which included a vague experience of a hen in the peripheries of vision. There is no free-floating, persistent hen ‘sense-datum’ independent of our perceptual experiences.
ways of “seeing that” which might be part of the perceptual experience but which require antecedently having linguistic concepts, such as, maybe, seeing a carburettor or hearing Mozart. And where does semantics itself fit in? After all, it feels different to read or hear something in a language one understands than in an alien tongue.

5) More evanescent components to sensations, such as a sense of familiarity (as with my house, compared with another similar one down the street) or knowing at a glance what is going on (when noticing an argument over change at a supermarket checkout for example).

6) Propositional attitudes like “believing that p” or “fearing that p.” Standardly, characterisations of the propositional attitudes have had precious little to do with conscious experience, and certainly beliefs and desires can persist unconsciously—one is not always aware that one believes Paris is in Texas, for example, but it is the same belief whether or not it is at the forefront of your mind; presumably, therefore, one could come to have a propositional attitude towards some description of an experience without changing the feel of that experience.

7) The feeling of bringing a propositional attitude to mind; presumably something changes between when we (unconsciously) happen to believe Paris is in Texas and when we bring that belief to mind—when we become aware of it. In what does this difference consist? What is the phenomenal property that is added to the belief to make it felt—perhaps the aural (or oral) experience of ‘talking to oneself’ plus a ‘feeling of assertion’ (rather than, say, denial)?

I am not asserting definitively that all these different kinds or components of awareness are discriminable and need to be taken account of by a complete theory; it seems to me an open question whether, for example, “seeing that” in some sense requires language use, or is perhaps completely independent of language and rests solely upon an ability to perceptually discriminate in certain ways. I suggest that consideration of these issues is at least partially an empirical
one\textsuperscript{24} because it seems to me inadequate to rely solely upon the resources of human introspection in these matters. It might be, for example, that experimental evidence could convincingly show that non-language-users possess what can only be called perceptual concepts, and that this conceptualisation dominates at such early stages of perceptual processing that it must be part of the consequent experience, and not some function upon that experience. Or it might be that future scientific accounts of the brain demonstrate that experiences can properly be called entities in the brain—fairly definable sets of properties of some more or less diffuse collection of neurones—whereas propositional attitudes are only virtual entities: behavioural consequences, perhaps, of highly complex network properties, such that unconscious beliefs and desires are not ‘there’ at all … all that the brain ‘contains’ is a holistic disposition to make current certain beliefs under certain conditions—to, perhaps, ‘tell yourself’ about them—and a set of constraints on behaviour. In such a case, it might again be most plausible to think of ‘experiencing that’ as being part of sensory experience, rather than the creation of an unexperienced propositional attitude modifying some basic experience of colours or sounds.\textsuperscript{25}

3. BEING UNINFERRED

Let’s return to the discussion of the directness of qualia and experiences (the latter still under-

\textsuperscript{24}There are also non-empirical considerations, such as the adequacy of externalist theories of content: although changes to the world outside our heads might make a difference to the propositional content of our experiences—whether we perceive water, $\text{H}_2\text{O}$ or water\textsubscript{2} (XYZ), for example—our experiences seem to be unaffected. As Seager says, “it seems evident that both I and my doppelgänger on Twin-earth will have the same experiences in some robust sense of ‘same,’ as we pursue our duplicate paths through almost duplicate worlds. Certainly we could be instantaneously exchanged back and forth with no one, including ourselves, being the wiser” (1991, 134).

\textsuperscript{25}This latter model was held by several sense-datum theorists; Price (1950, 8ff.), for example, argued that “perceiving that” involved a relation between some basic experience and an inference from, description of, or attitude towards those coloured shapes or uninterpreted phonemes. C.I. Lewis, similarly, held that “the two elements to be distinguished in knowledge are the concept … and the sensuously given,” that “knowledge of the objective arises [only] through conceptual interpretation of the given,” and hence that “there is no knowledge merely by direct awareness” (1929, 37).
stood in my semi-technical sense). The third variety of directness that can be isolated from the literature—and that which was probably the most common and influential—was that of being uninferred. For example, H.H. Price wrote that, undoubtedly, when seeing a tomato, there exists a red patch of a round and somewhat bulgy shape, standing out from a background of other colour patches, and having a certain visual depth, and … this whole field of colour is directly present to my consciousness. … And when I say that it is ‘directly’ present to my consciousness, I mean that my consciousness of it is not reached by inference, nor by any other intellectual process (such as abstraction or intuitive induction), nor by any passage from sign to significate. … This peculiar and ultimate manner of being present to consciousness is called being given, and that which is thus present is called a datum. The corresponding mental attitude is called acquaintance, intuitive apprehension, or sometimes having. Data of this special sort are called sense-data. (1950, 3)

Price, therefore, defines “givenness” negatively as being such that one’s apprehension of the given is reached in some way other than by:

a) any intellectual process analogous to inference or abstraction, or

b) the process of moving from a sign to its meaning or signification.

By condition a) I take it that Price means we are not acquainted with sense-data through the following sort of process: that of moving from some distinct thing that might count as some kind of evidence for their existence and qualities to the postulation (no matter how well-justified) of their existence and qualities. Thus our apprehension of sense-data is not like our apprehension that \( Q \) must be true when we already know that \( P \supset Q \) and \( P \) are true. Nor is it like our awareness that it will soon come about that \( Q \) because, in the past, nearly every time situation \( P \) has held it has been closely followed by \( Q \). Nor is it like our awareness that something is a member of the class \( Q \) because among its many properties \( \{F_1, F_2, F_3 \ldots\} \) are some that mark it out as a member of that class.

By b) Price means that sense-data are not the significates of some sign, either natural or conventional. They may themselves be signs of some sort, but our apprehension of them is not mediated by a sign, by contrast with, say, the ideas in Price’s book (which are signified by the words on the page) or the fact that one’s house is on fire in a situation where one sees smoke.
pouring out of the windows (where the smoke is a signifier and not the fire itself).

What shall we say about this kind of claim about the directness of qualia? It depends whether we are concerned with phenomenal or propositional awareness. In the former case, claims like Price’s are eminently plausible; qualia are, by their very nature, phenomenal and this, as we said above, is what constitutes their being apprehended phenomenally. We do not move from our apprehension of something else to our awareness that we are experiencing a visual sensation of a green lawn—the sensation of the green lawn is itself the phenomenal apprehension of that bundle of qualia. This, I think, is the important kernel of truth in the historical advocation of the directness of our apprehension of experience.

What about propositional awareness, however? Here Price’s thesis is much less clear cut … and it is with respect to this kind of “knowledge”—that of facts or truths—that the directness assumption has been most often attacked in the literature (e.g. Hirst 1959, 47–51). The problem here is that we are now dealing with two mental ‘objects’: the experience itself ($Q$), and a description of that experience ($P$), and it might well be that the latter is “inferred from” or “signifies” the former—if some words on a page signify an actual piece of scenery, for example, then why should we not think that analogously a mental proposition is a sign for the experience it describes?

Further, since, in the strictest sense, it is usually thought to be $P$ (the proposition), and not $Q$ (the quale itself), which is “believed” or “known” or “doubted,” and so on, it appears that it might be correct in one way to say that our knowledge of qualia is indirect in the sense we are discussing here. That is, one might say, we can talk loosely of direct apprehension of qualia,

---

26 One might feel the urge to talk about, say, the “movement” from a pattern of retinal stimulation to the phenomenal awareness of the lawn, and suggest that this is a kind of process of inference: if this is part of what Price means, then we are back to what I just discussed as directness as being uninterpreted, and the thesis that qualia are given loses plausibility accordingly. For example, the influential researcher into perception Irvin Rock defended what he called an “indirect” account of perception, which he described as the claim that sensory experience is derived by inference (Rock 1996); however, by this he apparently meant that sensory experience is not wholly determined by environmental stimulation (retinal patterns, for example), and as such comes closer to the view discussed in the previous section: that sensory experience is not uninterpreted.
when we mean phenomenal awareness, but really our knowledge of qualia is indirect in the sense of being inferential. I do not actually endorse this position, and intend as far as possible to remain neutral on it. Clearly, it relies upon a particular notion of “knowledge” which explicitly operates only over linguistic entities, and I do not know such an account to be ultimately supportable. Also, it might still be that $P$ is psychologically related to $Q$ in some way other than by inference or signification: this was a line of argument pursued fairly extensively in the heyday of the sense-datum debate, and three distinct arguments to this effect can be identified—all, I think, inadequate.

a) The Regress Argument

The first argument for the claim that propositional awareness of qualia is non-inferential is just that some $P$ must be directly present to consciousness in Price’s sense, or we should be in an infinite regress.

The subject or subject-matter about which we think must be somehow brought before the mind, if we are to think about it, and it cannot always be brought there by previous thinking, or we should have an infinite regress. This means that something must be given. And sensing is one of the ways (I do not say the only one) in which subject-matters for thought are given to us. (Price 1950, 7)

Price does not expand on this, but I take him to mean that, if we make something like inferences at all, then we must make them from some (propositional) starting point. If this starting point is itself the result of an inference, then there must be some earlier fulcrum for this earlier inference, and so on ad infinitum unless some kind of epistemological bedrock is reached. Similarly, if anything is signified by a sign, and if that sign itself is signified by some other sign, and so on, then unless at some point there is a sign that is known without the mediation of another sign the process will continue indefinitely.

This argument, however, is a weak one. First, it seems to assume that inference or signification

---

27 Or at least, it relies upon some account of beliefs and desires which requires them to be different entities than the experiences they are about.
tion are unidirectional and linear, forming great chains, whereas some would claim that this need not be so. Price apparently assumes that if $B$ is derivable from $A$, and $C$ is derivable from $B$, then $A$ cannot legitimately be derived from $C$; the holist, however, need not accept this.

Second, even if Price could establish that, for any finite system of knowledge or of signs there must be uninferred propositions or unmediated signs, it would remain to be shown that these uninferred representations are propositions about qualia; he relies upon the implicit undefended premise that, if anything is given, beliefs about experience are. Finally, formulated in this way, the argument misses its target because it does not take into account the possibility that all propositions are inferred but that the starting point for such inferences is experience (rather than some other proposition): that is, it assumes that $P$ is not inferred from $Q$, and so begs the question.

\begin{quote}
b) Epistemological Analysis
\end{quote}

Bertrand Russell, another pioneer of sense-datum analysis, at one time (1914) argued from a slightly different angle that propositional apprehension of qualia is uninferred. Instead of examining knowledge of ‘sense-data’ to see if it is given, he moved from a critical analysis of the epistemological status of our “common knowledge” to the claim that common knowledge is, as a matter of fact, divided into the derivative and the primitive, and then argued, on the basis of this critical analysis, that certain knowledge given by the senses is of the latter sort: “[T]he immediate facts perceived by sight or touch or hearing do not need to be proved by argument, but are completely self-evident” (1914, 75). Since Russell’s usage of the term “primitive” is roughly synonymous with the notion of “directness” as uninferred-ness that we are interested in here, this constitutes an argument that at least some (propositional) knowl-

\begin{footnotesize}
\begin{enumerate}
\item Though, he writes, “psychologists … have made us aware that what is actually given in sense is a lot less than most people would naturally suppose, and much of what at first sight seems to be given is really inferred” (1914, 75).
\item Russell distinguishes between the “logically primitive” and the “psychologically primitive.” It is the latter with which we are concerned here. To be “logically primitive,” for Russell, is to be such that “it is not the result of any
\end{enumerate}
\end{footnotesize}
edge of qualia is direct.

The major drawback of this analysis of Russell’s is that it is primarily \textit{a posteriori} and empirical, as Russell recognises,\textsuperscript{30} and is as such defeasible by new evidence (or new examination of the old). Worse, the experimental empirical evidence Russell himself gives to back up his claims is distinctly underwhelming—in fact, non-existent. What he mostly appears to be relying upon is psychological evidence to the effect that “psychological primitives” cannot be \textit{doubted}—they require no further justification than their merely being believed. He calls such apprehensions—“those which resist the solvent influence of critical reflection” (1914, 77)—“hard data,” and says that making up the bulk of “the hardest of hard data” are “the particular

\textsuperscript{30} For example, he notes (1914) that distinguishing between what is really “given” in sense and what is inferred is neither an easy nor an \textit{a priori} task, but rather one involving empirical psychological analysis in place of phenomenological introspection.
facts of sense” (77–78).31

The more we reflect upon these, the more we realise exactly what they are, and exactly what a doubt concerning them really means, the more luminously certain do they become. Verbal doubt concerning these is possible, but verbal doubt may occur when what is nominally being doubted is not really in our thoughts, and only words are present to our minds. Real doubt … would, I think, be pathological. (1914, 77)

This argument from epistemological analysis, then, either relies upon putting forward detailed and relatively complete empirical evidence for the absence of an inferential link between experiences and propositional awareness of that experience—evidence which, I believe, is not yet available—or rests upon evidence for the psychological certainty of propositional knowledge and somehow moves from there to a claim of directness. Let us, finally, consider this second option in a little more detail.

c) Certainty

This third kind of argument for the givenness of propositional knowledge of experience is the most commonly encountered. It typically consists just in claims to the effect that knowledge (propositional awareness) of experiences is in some way indubitable or certain, and (therefore) uninferred.32 Thus we have Russell’s argument explored above and Price’s already quoted move from the certainty “that something is red and round then and there” to its being “directly’ present to my consciousness” (1950, 3). Norman Malcolm even went so far as to define the term in

31 At this time Russell considered also “the general truths of logic” to be hard data—these are not themselves psychologically primitive but include beliefs deductively proved from psychologically primitive premises. Other, more subsidiary, varieties of hard data for Russell included some facts of memory and introspection.

32 This appeal to the Cartesian method of doubt (although few seem to call it this or indeed acknowledge that this is what it is) seems to be the most basic argument in the literature for the givenness of sense-data. That is, if one doubts everything until one reaches something that cannot be doubted, one is left with, at least, sense-data. So, for example, I could withhold assent from the claim that there really is an envelope there, that it really is white, that it really has the shape it does, and so on, but it is impossible for me to doubt that it appears to be a rectangular, white, envelope. Or rather, since perhaps I could raise sceptical doubts about my use of the concepts involved, it is better to say that I cannot withhold assent from the claim that my ‘visual field’ appears in just the way it does.
this way: “A directly perceives x if and only if A’s assertion that he perceives x could not be mistaken” (1963, 89).

By itself, clearly, this argument is incomplete: it is not at all obvious, prima facie, that some proposition could not be at the same time indubitable and the result of an inference … indeed, several logical truths spring to mind immediately as a possible counter-example. What missing premise could be supplied which might provide a reason to think that certainty is good evidence for directness? One possibility would be to argue that the Method of Doubt is precisely the bringing into question of inference and reasons for belief: this would suggest that the absence of possible doubt indicates the absence of inference. Thus, one might argue, logical truths can be doubted as they are the result of a chain of inference and the validity of that inference can be doubted; propositional awareness of experience, by contrast, cannot be doubted; if such beliefs were the result of inference they could be doubted in the same way as logical truths; they are, therefore, uninferred, i.e. “directly apprehended.”

Though much more plausible than its unreconstructed version, this argument from certainty is still fairly unimpressive. In particular, it is not clear that propositional awareness of experience really is “certain.” For example, since the description of an experience is, confessedly, a different mental entity than the experience itself, it is clearly at least logically possible that the former can occur without the occurrence of the latter—hence, logically, one might sincerely but falsely believe (propositionally) that one is undergoing a certain experience. Secondly, it is also logically possible that one might misdescribe one’s experience: one might believe one is having a chartreuse experience, for example, when really it is luteous.33 (I will examine the certainty of one’s awareness of qualia in much more detail in the next chapter.)

Further, that inferred beliefs may be doubted does not show that undoubted beliefs must not be inferred; this claim requires the rather stronger—and harder to defend—thesis that only inferred beliefs may be doubted. Worse, even granted that claim, that beliefs happen to be undoubted does not entail that they cannot under any circumstances be doubted—for example, it

---

33 Luteous = Light or moderate greenish yellow in colour.
might be that our propositional awareness of experience is actually inferred but that, for some reason, we cannot discover this fact and thus, as it happens, never doubt the results of such awareness. In other words, the psychological fact of absence of doubt, even under conditions of Cartesian rigour, still permits the possibility of indirectness.

4. CONCLUSION

So is our apprehension of qualia “direct,” or “immediate,” or not? In the first two senses of “direct”—“unmediated” and “raw”—it is phenomenal awareness that is of most interest; in the third sense—“being uninferrred”—it is propositional awareness to which most attention is usually paid. On the view defended in Chapters One through Three, qualia certainly (and almost trivially) are phenomenally apprehended without inference, and it makes good sense to speak of this apprehension as being “direct” in the sense of being unmediated. On the other hand, propositional apprehension of experience is probably not “direct” on any interesting interpretation of the word. This table summarises our results for this chapter:

<table>
<thead>
<tr>
<th></th>
<th>Phenomenal Apprehension</th>
<th>Propositional Apprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmediated?</td>
<td>Probably yes.</td>
<td>Depends on preferred usage.</td>
</tr>
</tbody>
</table>
Chapter 5: The Certainty of Qualia Apprehension

The second element of the epistemological “givenness” of qualia is the putative certainty of our apprehension of them. Before asking whether knowledge of qualia is indeed “certain,” it is worth getting clearer about just what this term is supposed to mean in this context. Though the core meaning here is clearly something like “without doubt” or “indisputable” the term can be used in this context (and has been so used, at one time and another) to refer to any one or more of a bundle of properties:

1) **Indubitability:** if \( A \) thinks of \( p \), \( A \) believes \( p \).
2) **Strong indubitability:** if \( A \) thinks of \( p \), \( A \) knows \( p \).
3) **Incorrigibility:** if \( A \) believes \( p \), \( p \) is true.
4) **Being self-verifying:** \( p \) is contingent, but if it is asserted sincerely it is asserted truly.
5) **Being evident:** if \( p \) is true, \( A \) believes \( p \).
6) **Being strongly evident:** if \( p \) is true, \( A \) knows \( p \).
7) **Solidity:** if \( A \) believes \( p \), \( A \) knows \( p \).
8) **Transparency:** \( A \) believes \( p \) if and only if \( p \) is true.

Even more confusingly, each of these conditionals can have more or less modal force and greater or lesser scope. Consider, for example, cases where \( p \) is “Daniel Dennett is not a zombie.” The strong indubitability of this claim might be restricted to just Dennett, or might include a wider group of people (such as all the people who have met Dennett); also, one might want to say that it is necessarily strongly indubitable for Dennett, or prefer to make a modally weaker claim. Further, the conditional claim itself might or might not be believed or known by

---

1 Williams 1978, Appendix 1, has an extremely useful breakdown of this material.
2 Williams calls this “irresistibility.”
3 Armstrong 1968a calls this being “self-intimating for \( A \).”
4 Armstrong (1968a, 101) discusses this. Obviously, this property is implied by being both incorrigible and evident.
5 That is, the force of the conditional might be material, formal or logical implication (and perhaps others as well). There may also be some room for modal qualification of the consequent. For example, possibly one might
\( A \) (or others), with more or less certainty: thus, for example, \( p \) might be incorrigible, but \( A \) might not know that it is. That is, the first order state of something’s being “certain” to one should not be conflated with the second order state of believing that something is “certain.”

Finally, there is a problem about what might be meant by “\( A \) knows \( p \)” in some of the consequents of these conditionals—in particular, introducing a hint of circularity, over to what degree belief in \( p \) must be justified in order that \( p \) can be said to be “known” … that is, what degree of certainty about \( p \) is necessary! I have no intention at this point of joining, let alone resolving, the debate on the proper definition of “knowledge”—feel free, therefore, to fill the gap here with your own preferred account of epistemic justification.

In this chapter I shall discuss in more detail the notions of indubitability, evidence and incorrigibility, not because I think the eight versions of certainty can be collapsed to these three core theses but just because I take them to be the most important and interesting properties in this domain.

Notice, first, that each of the definitions of these notions involve some proposition \( p \), which can be the object of belief or disbelief, and which is either true or false: we are chiefly concerned in this section, then, with propositional, rather than phenomenal, awareness of qualia. Insofar as there are any analogue notions for phenomenal awareness, we can make short work of the discussion; the following are probably the best candidates for phenomenal versions of the three kinds of certainty we are most interested in:

9) *Phenomenal indubitability:* if \( A \) experiences \( q \), \( A \) phenomenally apprehends \( q \).

10) *Phenomenal evidence:* if \( q \) is a certain way, \( A \) phenomenally apprehends \( q \) as being that way.

11) *Phenomenal incorrigibility:* if \( A \) phenomenally apprehends \( q \) as being a certain way, \( q \) is experienced in that way.

In these definitions \( q \) is a quale or experience, rather than a proposition. Now, the characterisation of qualia we are using here, on the basis of the considerations in Chapter Three, effectively collapses any distinction between instantiating or undergoing a certain experience and

want to make the (slightly odd) claim that if \( A \) thinks of \( p \), then \( A \) necessarily believes \( p \).
phenomenally apprehending that experience—on this account, all there is is the experience. Seeing a red ball involves instantiating certain phenomenal properties, such as ‘redness’ and ‘roundness,’ and the very instantiation of these properties is their phenomenal apprehension—the awareness of them. Each of these three conditionals, therefore, is just trivially true; in this sense, the esse of qualia is their perciπi—or, better (though still a little unsatisfactory), sentiri. Thus, phenomenally speaking, qualia are necessarily indubitable, evident, and incorrigible.6 As C.I. Lewis once put it, “[a]pprehension of the presented quale, being immediate, stands in no need of verification; it is impossible to be mistaken about it. Awareness of it is not judgement in any sense in which judgement may be verified; it is not knowledge in any sense in which ‘knowledge’ connotes the opposite of error” (1929, 125).

Now, on to the indubitability, evidence and incorrigibility of propositional apprehension of qualia.

1. INDUBITABILITY

Perhaps the most fundamental of the kinds of certainty is that typically meant by Descartes when he wrote of ideas being “indubitable” or “certain.” For Descartes, something is indubitable if the attentive mind “is unable not to assent” to it—something that, when it is “clearly and distinctly perceived,” it is impossible for the mind to doubt.8

…[T]he nature of my mind is such that I would be unable not to assent to these

---

6 William Seager correctly comments that “the apparent oddity of the doctrine of infallible access stems mostly from illicitly thinking of our awareness of our own modes of consciousness as a perceptual process with qualia on one side and conscious appreciation of them on the other (just like seeing). This perceptual model is not correct, is in fact misleading…” (1991, 144).

7 I think I can wholly endorse the sentiment behind this form of words, though I do not agree with everything Lewis has to say about qualia, even in the rest of this passage.

8 Again (though the distinction is not evident in Descartes, and thus I am probably simplifying the historical literature) we can see that we are talking here about the propositional apprehension of qualia—apprehension that some fact to do with qualia is the case, that some proposition is true of some quale—rather than their phenomenal apprehension.
things (which I clearly and distinctly perceive) so long as I clearly perceive them. (1911, Vol. I, 180)

I shall take “clearly and distinctly perceived” here to pick out those situations where something like the maximum of critical and analytical attention is being devoted to an idea. (Bernard Williams, in his lucid and careful book Descartes, also adopts this usage, though he later concludes that Descartes himself often uses the phrase, rather circularly, as if it means to see that an idea is transparently true.)

The basic idea here is that it is impossible to will to doubt that which is both certain in this sense and clearly and distinctly perceived.

If there is any certainty … it must concern those things which are clearly perceived by the intellect. But of these some are so evident and at the same time so simple that we can never think them without believing them to be true…. For we cannot doubt them, unless we think of them; but we cannot think of them without at the same time believing them to be true, as has just been laid down; therefore we can never doubt them without at the same time believing them to be true; that is, we can never doubt them. (1911, Vol. II, 42)

However, the indubitability thesis is a more limited one than it might at first seem. That \( p \) is indubitable does not entail any of the following claims:

a) that \( p \) is in fact true;
b) that \( p \) cannot be doubted at all;
c) that \( p \) cannot be doubted even by the subject who finds \( p \) indubitable;
d) that the second-order claim “\( p \) is indubitable” cannot be doubted at all;
e) that the second-order claim “\( p \) is indubitable” cannot be doubted even by the subject who finds \( p \) indubitable.

That a) fails to follow is made clear by the fact that it is not inconsistent to assert both that \( A \) finds \( p \) indubitable and that \( p \) is false. Perhaps Euclid’s “parallel postulate” might serve as an example here: for several centuries it was universally held to be indubitable, yet it turned out to

---

9 See Williams 1978, 182–183.
10 Williams makes similar points, though he does not go into great detail, and does not distinguish between b) and c) and d) and e).
be false. As one author has put it, “[o]nce Lobachevskian geometry is known it is hard to argue, as Kant did before it was known, that Euclidean propositions are true of our \textit{a priori} intuitions” (DeLong 1970, 49). Less radical examples are also available: thus, I might for some reason be unable to doubt that I have a guardian angel, though guardian angels do not exist.\textsuperscript{11}

A \textit{caveat}, however, needs to be entered here, using an argument about knowledge that parallels that of Descartes’ doubt above. It does seem that it \textit{would} be impossible, and perhaps even logically incoherent, for someone to assert “I find $p$ indubitable and I believe that $p$ is false,” since that amounts to saying “I believe $p$ and I do not believe $p.$” Thus if $p$ is indubitable for me then, by definition, at least while I am considering $p$ I must believe $p$; if I believe $p$ then it seems reasonable to say that I cannot sincerely hold $p$ to be false; if I cannot believe $p$ to be false then I cannot fulfil a minimum condition for knowing $p$ to be false; thus as long as $p$ is indubitable to agent $A$ then $p$ \textit{cannot} be known to be false by agent $A$.

On the other hand, it is not so clearly inconsistent for $A$ to \textit{entertain the possibility} that $p$ is false. This may be so for three reasons. Firstly, the property of indubitability, for Descartes at least, if it holds for some $p$, leads inevitably to belief \textit{only at the time that $p$ is being considered}—and not just casually considered but “clearly and distinctly perceived,” scrutinised with maximal critical and analytic attention.\textsuperscript{12} Thus, even indubitable ideas can be doubted by those who do not attend to them properly such as, in extreme cases, where $p$ is really only mentioned and not used. For example, it is possible for someone to sincerely assert that “I do not believe all triangles have three sides,” especially if they are not thinking properly about what it is to be a triangle, or do not even know what the word means—nevertheless, we might want to say that “triangles always have three sides” is an indubitable proposition. Similarly, even someone who understands the concept \textit{triangle} might sometimes, prior to reflection, begin to doubt that the

\textsuperscript{11} The first example works for whether $A$ can rightly count $p$ indubitable (a second order judgement), and the second for whether $A$ can doubt it (a first order comment).

\textsuperscript{12} That, for Descartes, something can be indubitable only when thought about clearly is actually exemplified by his own process of Pure Enquiry, for at first he \textit{rejects most of} those ideas which are indubitable: “… [T]he Doubt got as far as it did only by a measure of inattention” (Williams 1978, 186).
definition “the plane figure formed by connecting three points not in a straight line by straight line segments” is indubitable. This kind of reasoning also shows why b), c), d) and e) do not follow from the fact that \( p \) is indubitable.

The second reason why \( A \) might entertain the possibility that \( p \) is false even though \( p \) is indubitable is because \( A \) might hold (correctly) that, while she cannot doubt \( p \), this does not by itself make it the case that \( p \) must be true. Thus, although she believes \( p \), and although she can envisage no situation in which she could think carefully of \( p \) and not believe \( p \) (and so not necessarily hold \( p \) to be true), nevertheless she recognises in some abstract way that \( p \) may not be true. For example, just as it seemed indubitable to nineteenth century mathematicians that two parallel lines must both run at 90° to the ‘perpendicular’ (although that is in fact false), it seems to me now indubitable that any sufficiently strong formal system of arithmetic is incomplete if it is consistent (Gödel’s first incompleteness theorem), but I recognise the abstract possibility that this claim is false … though not that I might have epistemic reasons to even take this possibility at all seriously—or even that anybody might ever have such reasons. It might be rather like recognising such unlikely and untestable possibilities as that an advanced, invisible alien race is observing Earth from space or that there exist entities which have no causal interactions at all with the rest of the universe.\(^{13}\)

\(^{13}\) At least two suggestive pieces of reasoning can be given in additional support of this. Firstly, it seems clear that most of our other beliefs can operate in this way: that is, although I believe \( p \) and therefore hold \( p \) to be true, I recognise the possibility that \( p \) might really be false. For example, I believe (it to be true) that Shakespeare was the author of *The Tempest*, but I would be willing to consider—and even to seriously entertain—arguments to the effect that he was not. Secondly, the following picture (adapted from a not unfamiliar one in the literature on scepticism) might have heuristic value. Consider a situation in which \( A \) has a set \( \Gamma \) of various indubitable beliefs that propositions \( p_1, p_2, p_3, \ldots \) are true. Each one of those beliefs is such that, by Descartes’ argument above, she cannot reject it or even call it into question: whenever she thinks clearly of the proposition it expresses, she comes to believe that proposition, and if she believes \( p_n \) she cannot simultaneously doubt it. However, she has the additional belief that \( q \)—the proposition that one or more of the members of \( \Gamma \) may be in fact false. She can even think carefully about \( q \) at roughly the same time as she is thinking clearly and distinctly about some or all of the members of \( \Gamma \) and still hold all those beliefs together: in such a situation \( A \) is neither psychologically nor logically inconsistent.
Chapter 5: “The Certainty of Qualia Apprehension.” 117

The third reason someone might give for denying the claim that \( A \) must hold \( p \) to be true if \( p \) is indubitable for \( A \) is that \( A \) might doubt that \( p \) is in fact indubitable: that is, it might occur to \( A \) that \( p \) only seems to be indubitable but in fact is not. Thus my parallel postulate example shows that Euclid’s fifth postulate was not, in fact, indubitable—it was just apparently indubitable for a time. Sufficient examination shows that it is in fact possible to entertain the idea clearly and distinctly and yet still hold it to be false—and presumably, it always was possible, though nobody actually did so. A similar argument can be made for any idea that is eventually doubted as long as that doubt persists in the face of—or, even better, is based upon—clear and critical examination of the idea.

If this series of arguments is allowed to be persuasive, then it seems that only two conditions will make a proposition truly indubitable: either it is in fact demonstrably true beyond any possibility of deception (whatever exactly this means); or it is beyond the psychological capacity of human beings to even take seriously the possibility that the proposition is false, though the proposition may as a matter of fact be false.

It follows from all this that not only is it not inconsistent to assert both that \( A \) (or some or all \( A \)) finds \( p \) indubitable and that \( p \) is false, but it is also not inconsistent to assert any of the following statements:

- “I find \( p \) indubitable, but \( p \) may be false.”
- “I find \( p \) indubitable, but I believe that it is, in some abstract sense, possible that \( p \) is false.”
- “\( p \) is indubitable to \( A \) but \( A \) believes that \( p \) is false.”
- “I find \( p \) indubitable, but it may not in fact be indubitable.”

Clauses b), c), d) and e) above all fail to follow from the property of indubitability (that is, they need not be true of propositions which are indubitable) in part because of the reasons given above. In addition, clauses b) and d) fail to follow from indubitability because it is not inherent in the notion that an idea which is indubitable to one person must be indubitable to all: that is, indubitability might be relativised to a particular subject. Thus, idea \( p \) might be indubitable for \( A \) but not for \( B \). Euclid, presumably, found the parallel postulate indubitable, but G.F.B. Ric-
We are now in a position to map out in some detail the “conceptual space” of the notion of indubitability. The important axes in this space seem to be:

i) the degree of attention which A gives to the idea of p;

ii) the range of subjects for which p is indubitable; and

iii) the modality of that indubitability.

It seems to me that three points are of especial salience along axis i): “merely entertaining” an idea (E), examining an idea as critically and analytically as the subject is able to (S); and examining an idea as critically and analytically as possible (P)—that is, as any (human) subject could. On axis ii) the two most important points are all subjects (∀), and some (at least one) subjects (∃). Along iii), the two significant modalities are something like “as a matter of contingent fact always” (⊃) and “necessarily always” (where this means simply “not possibly not”). In addition, we have seen, there are two ways in which p could be necessarily indubitable: if it is “very obviously true” (7) and if it is impossible for the subject to doubt it (though it may be false) (⊥).

Thus we can distinguish the following eighteen (!) varieties of indubitability:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>(For any subject, (if A entertains the idea of p, then A believes p)) and p may be false.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>∀</td>
<td>E</td>
<td>⊃</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(For any subject, necessarily (if A entertains the idea of p, then A believes p)) and p is true.</td>
</tr>
<tr>
<td>2</td>
<td>∀</td>
<td>E</td>
<td>⊤</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(For any subject, necessarily (if A entertains the idea of p, then A believes p)) and p may be false.</td>
</tr>
<tr>
<td>3</td>
<td>∀</td>
<td>E</td>
<td>⊥</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(For any subject, (if A thinks as critically and analytically as A is able about the idea of p, then A believes p)) and p may be false.</td>
</tr>
<tr>
<td>4</td>
<td>∀</td>
<td>S</td>
<td>⊃</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(For any subject, necessarily (if A thinks as maximally critically and analytically as A is able about the idea of p, then A believes p)) and p is true.</td>
</tr>
<tr>
<td>5</td>
<td>∀</td>
<td>S</td>
<td>⊤</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(For any subject, necessarily (if A thinks as maximally critically and analytically as A is able about the idea of p, then A believes p)) and p</td>
</tr>
<tr>
<td>6</td>
<td>∀</td>
<td>S</td>
<td>⊥</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>∀</td>
<td>P ⊃ (For any subject, (if A thinks as maximally critically and analytically as any subject is able about the idea of p, then A believes p)) and p may be false.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>∀</td>
<td>P ⊤ (For any subject, necessarily (if A thinks as maximally critically and analytically as any subject is able about the idea of p, then A believes p)) and p is true.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>∀</td>
<td>P ⊥ (For any subject, necessarily (if A thinks as maximally critically and analytically as any subject is able about the idea of p, then A believes p)) and p may be false.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>∃</td>
<td>E ⊃ (For some subject, (if A entertains the idea of p, then A believes p)) and p may be false.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>∃</td>
<td>E ⊤ (For some subject, necessarily (if A entertains the idea of p, then A believes p)) and p is true.</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>∃</td>
<td>E ⊥ (For some subject, necessarily (if A entertains the idea of p, then A believes p)) and p may be false.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>∃</td>
<td>S ⊃ (For some subject, (if A thinks as critically and analytically as A is able about the idea of p, then A believes p)) and p may be false.</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>∃</td>
<td>S ⊤ (For some subject, necessarily (if A thinks as maximally critically and analytically as A is able about the idea of p, then A believes p)) and p is true.</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>∃</td>
<td>S ⊥ (For some subject, necessarily (if A thinks as maximally critically and analytically as A is able about the idea of p, then A believes p)) and p may be false.</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>∃</td>
<td>P ⊃ (For some subject, (if A thinks as maximally critically and analytically as any subject is able about the idea of p, then A believes p)) and p may be false.</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>∃</td>
<td>P ⊤ (For some subject, necessarily (if A thinks as maximally critically and analytically as any subject is able about the idea of p, then A believes p)) and p is true.</td>
<td></td>
</tr>
</tbody>
</table>
∃P

(For some subject, necessarily (if A thinks as maximally critically and analytically as any subject is able about the idea of p, then A believes p)) and p may be false.

So, then, now we are clearer about what it is to be indubitable, is the propositional apprehension of qualia indubitable in any of these eighteen ways? Disappointingly (as, we saw above, Russell found out) it is difficult to answer this question a priori, since it is at base an inquiry about the actual psychological capacities of human beings to withhold belief under certain conditions, and this—unlike, say, normative questions concerning truth or epistemic justification—is an almost wholly empirical question. What we should do in principle, to deal with this question definitively, is answer the following difficult questions about propositions about qualia:

1) to whom, if anybody, are they indubitable?
2) under what conditions of belief-scrutiny, if any, are they indubitable?
3) if they are indubitable, are they necessarily indubitable in some way?\(^{14}\)

Nor need we assume that all propositions about qualia are equal in these respects; (impossibly) ideally we should ask each of these three questions of each individual proposition about qualia.

Notwithstanding this rather depressing (from the standpoint of the philosopher, at least) conclusion, it might be worth making some, hopefully plausible, a priori observations. First of all, it seems likely that a large class of propositions ‘about’ experiences are not indubitable in most of the eighteen ways above even if they are in fact true. Roughly put, these are propositions that imply or assume the existence of entities or properties in addition to the qualia themselves: for example, “I am perceiving a table” is arguably not indubitable for most or all people since it can be doubted whether ‘the table’ actually exists, as opposed to merely appear-

\(^{14}\) It is rather difficult to see how question 3) could be answered without important extra pieces of knowledge: we would have to know with certainty either that \(p\) is true and that this level of certainty is almost involuntarily attainable by some or all people; or we would have to know a great deal about the absolute capacities of human beings with respect to withholding assent to believe.
On the other hand, it might be objected that “I see a table” could well be, in fact, indubitable in senses 10 or 13: all that would be required to satisfy these conditions would be that there is at least one agent who, as a matter of fact, always comes to believe that she sees a table whenever it is true and she entertains (or scrutinises) the idea. Indeed, perhaps most people who have not been exposed to sceptical methods and ideas (and who have never taken mind-altering substances) are like that. Further, perhaps there are at least some people who, unlike Descartes, no matter how carefully they critically examine their belief in tables, never once come to reject or question it: this would make such a proposition indubitable in at least sense 15. Perhaps the only secure claim in this area, if any are secure at all, is that beliefs like “I see a table,” when they are contingently true, are not indubitable in senses 4–9.

Second, when $p$ is not true it is, usually, unlikely to be indubitable. Though it is difficult to rule out a priori the possibility that someone would find that consideration of such propositions always leads to belief in them, such an agent would be a very odd one indeed: for her, merely to (let us say) think hard about the claim “I experience the visual sensation of a red table”—perhaps in bed at night in a darkened room—would bring it about that she believes that she does in fact sense a red table (even though she does not). Similarly, when perceiving a uniform green wall, surely no amount of careful consideration of the proposition that one is experiencing redness will reliably bring it about that one believes one is seeing red.\(^{15}\)

More generally, it seems hugely unlikely that, under normal conditions, for normal people, thinking about any particular quale (no matter under what description—i.e. even thought about in terms of colour patches and so on) can always make it the case that one believes one is currently experiencing that quale. On the whole, it seems reasonable to say, entertaining propositions about qualia is not part of the causal or cognitive networks that may give rise to current qualia themselves or to beliefs about them.

\(^{15}\) For more detail on this, see the discussion of the relationship between phenomenal and propositional awareness in the section on incorrigibility, below.
On the other hand, it is plausible to suggest that thinking carefully about what it means to have qualia (not necessarily just perceptual qualia) generally will reliably lead under normal conditions to the belief that one is currently experiencing qualia. That is, this general proposition is apparently indubitable in at least the senses 13 and 16 above, and perhaps 10 also. However, we cannot yet make this a universal or a necessary claim. Daniel Dennett, for example, claims that 13 is not true of him under normal conditions, and thus (if he is sincere) it is not true of every agent. And (possibly) it would not be true either of an agent immersed completely in a sensory-deprivation tank, and so, for any agent, it is not necessarily the case that if they think carefully about the proposition that they are currently experiencing qualia they will come to believe it.\textsuperscript{16}

Finally, we should consider particular true propositions to the effect that certain qualia are occurrent—claims like “I am currently experiencing the visual image of a red ball” or “the mu-

\textsuperscript{16} Thus, 10, 13 or 16 can only be true for agents who have not been in a sensory-deprivation tank. What about the claim that “I have at one time experienced qualia”? This might well be always believed by some people who are thinking about what it means to say this but not currently experiencing qualia. Hence, for those people, it might appear to be necessarily indubitable: that is, in senses 14, 15, 17, 18 and perhaps 11 and 12. But this is not so. As an extension of Descartes’ Method shows, it is logically possible to doubt this proposition under those circumstances, and thus possible not to believe it. For example, one could think carefully about what it means to have once experienced qualia but entertain the possibility that all one’s memories of qualia are illusory (implanted by a \textit{malin génie} perhaps), as is one’s understanding of the concept of qualia; thus, it is possible to withhold belief from the claim that “one has at one time experienced qualia.” Put another way, it is presumably logically possible that there could be an entity which understands conceptually what qualia are but has never experienced them; and further, it is sometimes psychologically possible for all human beings to entertain the possibility that they are such an entity; therefore, $p$ is not necessarily indubitable where $p$ is “I have at one time experienced qualia.”

On the other hand, someone might still plausibly object either that:

a) For some human beings, even if they were in a sensory-deprivation tank, it would not be psychologically possible for them to doubt that they once had qualia. That is, $p$ would be indubitable in senses 15 and perhaps 12.

b) For some or possibly all human beings, it is psychologically or logically impossible to have the concept of qualia without having once had qualia; further, and importantly, careful consideration of that concept shows this to be the case. Therefore, under careful examination, the relevant subjects would be unable not to realise that the possession of the concept entails that they once experienced qualia, and thus be unable to doubt that they once experienced qualia.
sic I am experiencing sounds discordant,” when these assertions are true. Probably, claims like these are usually indubitable for most people: when one both experiences a red sensation and considers the proposition that one is sensing red, it seems very unlikely that one could fail to come to believe the proposition. Even in these cases, though, it would not do to be too universal in this claim. Possibly, for example, someone in the grip of a theory might sincerely withhold belief from the proposition, on the grounds, perhaps, that they are not experiencing a red sensation but “experiencing redly” in the Strong Adverbialist sense (see Chapter Three, section four), or merely colourlessly discriminating objective redness.

Further, there is a good prospect that one might experience certain qualia but not always be able to form beliefs about them. It might be that certain qualia are too evanescent and fleeting to be captured by propositional awareness (even though they are, necessarily, phenomenally apprehended): stimuli on the threshold of subliminality, or on the very peripheries of sensory fields, might plausibly be experienced in this inherently non-propositional way. In such cases one could, in principle, carefully consider but withhold belief from true propositions like “I just saw the word ‘ice-cream’ on that slide” or “I just now sensed movement out of the corner of my eye.” Further, it seems that even within experiences that are generally attended to, there may be details that are not noticed. Suppose, for example, that, on two separate occasions, very close together, you were in a particular room containing a small, peculiarly shaped statuette. On the first occasion you scanned the room, qualia brought about by the statuette were in your visual field (hence, phenomenally apprehended), but you did not notice it. Had you later been asked whether you had seen a statuette when you were in the room, you would have denied it: you did not believe you had seen it. But on the second occasion, with close to qualitatively identical qualia making up your visual field, a friend draws your attention to the statuette. Now you would presumably be prepared to assent to it. And this example would still work even if your friend had drawn your attention to the statuette by saying “Did you see that unusual statue of the Hindu god Ganesha over in that alcove,” to which you would have replied “Yes”—indeed, you were aware of it all along, but had no beliefs about it.

In summary, then, though this is at bottom an empirical matter, I suggest that if proposi-
tions about qualia are ever indubitable at all then, probably, this is only when the propositions in question are in fact true, and that, within this set, claims of indubitability are most reliable for qualia that are attended to, or for more general claims such as that “I am currently experiencing some perceptual qualia.”

2. EVIDENCE

We defined “being evident” above as being such that if \( p \) is true, then \( A \) believes \( p \). This, I think, is a simpler concept than that of indubitability, though again we must beware of variations in modality and scope of the conditional. In addition, a distinction between what we might call “conscious” (C) and “unconscious” (U) belief acquisition is useful here: simply, one acquires the belief that \( p \) “consciously” if one not only acquires the belief \( p \) but is aware that one now believes \( p \); whereas the same belief might be acquired “unconsciously” if one simply, say, acquires the disposition to assent to \( p \) without noticing that one has done so. So we have eight interesting varieties of being evident:

<table>
<thead>
<tr>
<th></th>
<th>( \forall )</th>
<th>C</th>
<th>( \supset )</th>
<th>For any subject, (if ( p ) is true, then ( A ) consciously believes ( p ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \forall )</td>
<td>C</td>
<td>( \Rightarrow )</td>
<td>For any subject, necessarily (if ( p ) is true, then ( A ) consciously believes ( p )).</td>
</tr>
<tr>
<td>2</td>
<td>( \forall )</td>
<td>U</td>
<td>( \supset )</td>
<td>For any subject, (if ( p ) is true, then ( A ) unconsciously believes ( p )).</td>
</tr>
<tr>
<td>3</td>
<td>( \forall )</td>
<td>U</td>
<td>( \Rightarrow )</td>
<td>For any subject, necessarily (if ( p ) is true, then ( A ) unconsciously believes ( p )).</td>
</tr>
<tr>
<td>4</td>
<td>( \exists )</td>
<td>C</td>
<td>( \supset )</td>
<td>For some subject, (if ( p ) is true, then ( A ) consciously believes ( p )).</td>
</tr>
<tr>
<td>5</td>
<td>( \exists )</td>
<td>C</td>
<td>( \Rightarrow )</td>
<td>For some subject, necessarily (if ( p ) is true, then ( A ) consciously believes ( p )).</td>
</tr>
<tr>
<td>6</td>
<td>( \exists )</td>
<td>U</td>
<td>( \supset )</td>
<td>For some subject, (if ( p ) is true, then ( A ) unconsciously believes ( p )).</td>
</tr>
<tr>
<td>7</td>
<td>( \exists )</td>
<td>U</td>
<td>( \Rightarrow )</td>
<td>For some subject, necessarily (if ( p ) is true, then ( A ) unconsciously believes ( p )).</td>
</tr>
</tbody>
</table>
Now, are propositions about qualia evident? The first thing to notice is that for any particular quale or experience there are presumably a very large number of true propositions about that quale or experience, and it seems clear that, at least, not all of those propositions are evident. Thus, for example, some experience will be instanced by a qualifier (or set of qualifiers) which has a certain spatio-temporal location, many other relational properties, and probably numerous intrinsic properties (such as mass, charge, chemical composition, and so); further, the experience itself will have plentiful relational properties (such as being phenomenally ‘noisier than’ another similar experience) and the qualia that make it up will typically be describable in lots of different ways. For example, suppose one had the experience of a glaucous colour-patch (say, an unripe kumquat), and that one attended to this quale: it is not at all clear that one would reliably come to believe that “I sense a glaucous colour-patch in the shape of a kumquat.”

Further, as noted above, it seems plausible to say that, since certain parts of our experienced perceptual fields are unattended to, there may well be qualia which are experienced but about which we do not form beliefs. Since we form no beliefs at all about these qualia, no propositions about them can possibly be evident in the sense of "evident" which I have identified.

So, if not all propositions about qualia are evident, are there nevertheless some that are? It is plausible to suggest—and often has been suggested, in one way or another by philosophers ranging from Descartes to Ayer—that certain beliefs are reliably acquired, at least unconsciously, and at least by normal human language-users, as a result of the occurrence of experience. For example, if someone were perceiving a green lawn, it would be highly peculiar if they refused to assent to propositions like “I am currently experiencing a sensation of greenness” or “I appear to be seeing a lawn” or even just “I am undergoing perceptual experience.” However, specifying \textit{a priori} exactly \textit{which} propositions about experience are evident in this way is a daunting task—in particular since we are once again, it seems to me, in the domain of psychological fact rather than logical necessity. The traditional set of epistemologically privileged propositions—roughly, those descriptions of the intrinsic character of experience which rely upon no assumptions about anything else—is of little use to us here. In the first place, that set is too broad, presumably including as it does such propositions as that “this experience is glau-
cous” or “I am sensing plangent ‘noises’”—which may be true, but not always spontaneously believed, even unconsciously. In the second place, the set is too narrow: that we believe $p$ if it is true should not be confused with the Cartesian desire to believe $p$ only if it is undoubtedly true; thus, a variety of propositions are plausibly thought to be evident, under normal conditions, despite not being certainly true: central examples are beliefs to the effect that we perceive objective, external objects and events, such as giraffes, bassoons and murders. Many of the other ‘extrapolations’ that the human nervous system engages in would presumably, also, count as evident, at least unconsciously, including such propositional apprehensions as that railway tracks are really parallel, that the snow on a distant peak is cold, that the growl of a tiger signifies danger, and so on.

Are all of these claims of evidence contingent and empirically defeasible? Are there any propositions which can be shown to be necessarily—or even universally—evident? There are a variety of reasons to be cautious about this prospect. First, note that we must restrict any claims of this sort to language-users; any other kind of experiencer cannot, strictly speaking, come to believe a proposition at all (though, as discussed above in Chapter Four section two, plausibly there are analogues of belief, and so of being evident, which might apply for animals). Secondly, now we are in the business of finding propositions about experience that, if they are true, must be believed by any language-user: presumably, then, they must be expressible in sentences that any conceivable language-user could employ—and this is a very tall order. Consider even such basic propositions as “I experience a green patch”; even for the normal English speaker, the distinction between perception and sensation—essential for making a claim that cannot be doubted—may not be part of their linguistic tool kit; what prospects then for a colour-blind language community, or one which does not distinguish between green and blue, or one which perhaps does not even have a first-person pronoun? Thirdly, we must bear in mind

---

17 There is also a complicated bundle of issues relevant here as to how we come to learn language at all. Sellars (e.g. 1956), for example, influentially criticised the epistemological primacy of experience by arguing that language must be first learnt as a set of conceptual responses to external, objective objects, before it can be applied to the contents of experience. However, in my discussion I simply assume subjects are already in possession of some
that the neural tokening of the belief that \( p \) is (or at least is likely to be) a distinct entity from the neural tokening of the *experience* that \( p \) describes.\(^\text{18}\) Not only, therefore, is it certainly logically possible that the latter can occur without the former—that, for some \( p \), \( p \) is true but not believed—but it is very likely that at least once in all the history of language-use in all the universe, such a disconnection actually *will* occur.

There are good reasons to think, then, that though a wide variety of propositions about experience are probably usually evident to most ‘normal’ human adults, *no* species of propositional apprehension of experience is necessarily and universally evident.\(^\text{19}\)

### 3. INCORRIGIBILITY

Some proposition \( p \) is incorrigible if, if \( A \) believes \( p \), then \( p \) is true. Little analysis of this relation, I believe, is necessary for clarification: the following will suffice.

---

\(^{18}\) If this suddenly has a ring of implausibility, you may be thinking of a kind of “awareness that” that, putatively, is somehow *part* of perceptual experience—see the previous section.

\(^{19}\) In his paper “Why You Can’t Make a Computer that Feels Pain,” Dennett proposes the following counterexample to the evidence of propositional awareness of qualia. We speak of the anaesthetic nitrous oxide, he claims, as if it “doesn’t prevent the occurrence of pain at all, but merely makes one insensible to it when it does occur (as one can be rendered insensible to the occurrence of flashing lights by a good blindfold)” (1978a, 221). However, such examples as this can be explained without denying evidence. It could be that this drug, and other analgesics, actually do “prevent” the occurrence of the pain; that is, perhaps, though they leave most of the neurophysiological pathways intact (and so do not alter the bulk of the brain activity associated with pain), they do crucially alter the state that instances the *experience* of pain, such that the state no longer instances the quale pain but either not a quale at all or some other quale (one “like pain muffled in cotton wool” for example). Second, it could be that some analgesics work by causing us not to *attend* to our pain qualia; under such a circumstance, one would experience the pain as unattended, unnoticed pain and therefore (presumably) much less severely. Only thirdly, and rather frighteningly, might it be that certain analgesics leave our qualia unaffected but disconnect them from the cognitive and behavioural states that typically accompany them; thus, one might feel excruciating pain but fail to come *to believe* that one does, or *remember* that one did, and so on. Let us hope Dennett is wrong in his analysis!
Different theories based upon the putative incorrigibility of certain “fundamental” propositions describing qualia are, of course, strewn throughout the philosophy of this century, especially the empiricist-style foundationalist epistemologies. Thus we have the atomic propositions that feature in Wittgenstein’s *Tractatus* (1922); the “protocol statements” (*Protokollsätze*) of Neurath and the Vienna circle (see Hempel 1935); and Russell and Ayer’s “basic” propositions or statements\(^20\) (e.g. Russell 1940 and Ayer 1946). These are propositions that belong to that

… class of empirical propositions of which it is permissible to say that they can be verified conclusively. It is characteristic of these propositions … that they refer solely to the content of a single experience, and what may be said to verify them conclusively is the occurrence of the experience to which they uniquely refer. (Ayer 1946, 10)

Propositions of this kind are “incorrigible” in the following sense:

… [W]hat is meant by their being incorrigible is that it is impossible to be mistaken about them except in a verbal sense. In a verbal sense, indeed, it is always possible to misdescribe one’s own experience; but if one intends to do no more than record what is experienced without relating it to anything else, it is not possible to be factually mistaken; and the reason for this is that one is making no claim that any further fact could refute. (Ayer 1946, 10)\(^21\)

---

\(^{20}\) Ayer recognises that, since the verification principle is intended to be a criterion for meaningfulness, the term “proposition” is problematic, particularly since it:

- a) restricts the criterion to sentences which express a proposition, and
- b) is otiose, since such statements must already be meaningful (i.e. express something which is true or false).

Thus he introduces the technical term “statement,” by which he means any indicative sentence, where a sentence is “any form of words that is grammatically significant,” and furthermore any two sentences (presumably, though Ayer does not say so, as long as they are indicative) which are “mutually translatable” express the same statement (1946, 8).

\(^{21}\) Ayer asserts that this is a case of “nothing venture, nothing lose,” since the proposition can be known to be true. But he also calls it “nothing venture, nothing win,” because that knowledge is uninformative: it offers no
Chapter 5: “The Certainty of Qualia Apprehension.”

Sadly, however, the situation is not quite so straightforward as Ayer and Co. once presented it. Most importantly, as David Armstrong once pointed out, for the physicalist it cannot be the case that any propositions about qualia are in principle incorrigible. Given that such propositions “are caused, as immediately as possible, [only] by perceptive experiences” (Russell 1940, 137), and given that no causal link is in principle reliable, it therefore might be that the effect could occur without the cause, and so the proposition be believed falsely—for example, one might believe that one is undergoing a red experience when in fact it is green, due to a misstep in the causal chain from the green experience to the propositional belief. In other words, though the causal link between experience and propositional awareness of that experience might be far more reliable than that between external objects and perceptual experience, it is nevertheless still a causal link and subject to the same upsets and mischances as the latter.

Further, even if one has it in mind to somehow reject any causal account of the link between experience and propositional apprehension of that experience, it remains the case that the experience and the propositional attitude towards that experience are two distinct entities, as we have pointed out above: thus, no matter what one’s view of the link between the two of them, the latter could in principle exist without the former, and so be false.

more information than “what is already afforded by the occurrence of the relevant experience” (1946, 10).

22 Armstrong argued that, “on any plausible construal,” awareness of states of the brain entails “a self-scanning process in the brain.” If materialism is true this process must itself be a physical mechanism and therefore subject to the possibility of error (1968a, 102–103). Paul Churchland has made the same point (1979, 96), and I’m sure others have too.

23 I take “perceptive” here to be sensory rather then necessarily perceptual (which as we have seen is a success term): thus beliefs about qualia can be caused even by hallucinatory sensations.

24 Another argument against necessary incorrigibility is the following: “It is clear that if there is incorrigible knowledge of our own mental states, then it cannot apply to the past, but only to the present. Put the mental state even a fraction of a second in the past, and error becomes logically possible” (Armstrong 1968a, 104). This has two uncomfortable implications, Armstrong urges. The first is that knowledge which is logically certain (such as that I am in pain right now) in the space of a fraction of a second shifts and becomes merely empirically certain (such as that I was in pain a second ago). This seems implausible. The second is that our reports—and our beliefs—must extend temporally beyond the time of logical certainty (the “introspective instant”) and thus cannot be necessarily incorrigible (1968a, 104–106).
Chapter 5: “The Certainty of Qualia Apprehension.”

Suppose, however, that we adopt, if only for historical consistency’s sake, a wilfully verificationist theory of truth, such that some proposition is incorrigible if it is both true and cannot even in principle be shown to be false. Thus, for the logical positivist, one might say, basic statements are confirmed by experience and so true, and further no additional empirical evidence could, even in principle, show them to be false (when they are really true). Since this definition of incorrigibility relies upon one’s inability to empirically show that the belief is false, rather than simply the actual impossibility of its falsehood, it might allow us to admit the aforementioned possibility that some basic beliefs are false yet still assert the necessary incorrigibility of basic statements.

Naturally, this move relies upon the claim that basic statements could never in principle be shown to be false; what might motivate such an assertion? Typically, this is a logico-linguistic thesis which essentially argues that our language—or even, some ideal formal language—does not contain the linguistic resources necessary to hold basic statements false. Thus, for example, Hilary Putnam (1964) once argued that there is no notion in English for an appearance of an appearance of red, and he claimed that the reason was

...that any state which cannot be discriminated from “looks-red” counts as “looks-red” (under normal conditions of linguistic proficiency, absence of confusion, etc.). What this illustrates, of course, is that the “incorrigibility” of statements of the form “that looks red” is to be explained by an elucidation of the logical features of such discourse, and not by the metaphor of “direct” access. (66)

Armstrong, similarly, argued (1968a, 111–112) that it is part of what it means to make a sincere report of a seeming that the claim is true. As he puts it, if I sincerely have a seeming, then necessarily I do have a seeming—“sincere seemings” are just not the kind of proposition that can be false. Meanwhile Daniel Dennett, rather mockingly, recommends the friend of incorrigibil-

---

25 In “Minds and Machines” (1960), too, Putnam claimed (72–73) that “How do I know I have a pain?” is a deviant (i.e. logically odd; there is no statement that it could be used to make in a normal context) question, whereas “How do I know Smith has a pain?” is not. His analogue case was a Turing machine programmed to write “I am in state A’ whenever it is in state A. In such a case, Putnam urged: “Even if some accident causes the printing mechanism to print: ‘I am in state A’ when the machine is not in state A, there was not a ‘misconception’ (only, so to speak, a ‘verbal slip’)” (1960, 83).
ity treat qualia as logical constructs out of qualia judgements such that “a subject’s experience has the quale $F$ if and only if the subject judges his experience to have quale $F$. We can then treat such judgings as constitutive acts, in effect, bringing the quale into existence by the same kind of licence as novelists have to determine the hair colour of their characters by fiat. We do not ask how Dostoevski knows that Raskalnikov’s hair is light brown” (1988, 55).

However, these kinds of claims about the logical structure of “seemings” discourse fly wide of the mark we are aiming for. All that they establish, if they establish anything, is that if $A$ sincerely believes she experiences quale $q$ then it is true that $A$ believes she experiences quale $q$—whereas, of course, what the verificationist proponent of necessary incorrigibility needs is the claim that if $A$ believes she experiences quale $q$ then she really does experience $q$. That is, what sincere seemings necessitate is just that it is a seeming of that type—where the seeming in question is propositional, so is what is guaranteed; the phenomenal truth-maker, if any, for the content of the seeming is not touched. In fact, many of the advocates of this kind of linguistic thesis are at pains to note that the incorrigibility of sincere seemings does not entail the truth of any claims about qualia. Thus Dennett writes that constitutive qualia judgings bring into existence only something akin to “theorists’ fictions,” and certainly not “empirical facts in good standing,” while Armstrong insisted that, from the incorrigibility of reports of pain seemings, nothing follows about pains—in particular, it does not follow that I do have a pain now. Instructively, he compares “seeming” reports with the claim that “I doubt whether unicorns would beat centaurs over seven furlongs.” If a sincere report, it follows that I do have this doubt; nothing, however, follows about unicorns and centaurs.

Would, then, an adjusted logico-linguistic thesis be possible to the effect that it is part of the very structure of talk about propositional awareness of experience that those propositions must be true? The fact that various thoughtful commentators—Paul Churchland (1979), Daniel Dennett (1988), Michael Tye (1989), Steven White (1986)—have felt themselves able to deny the truth of basic beliefs about qualia without obvious inconsistency or misuse of language suggests that this prospect is not hopeful.

A different verificationist tack would be to argue that basic statements are necessarily em-
pirically privileged—that, for some reason, experiences are a class of objects that only the experiencer can know about, and that therefore there is no prospect of the experiencer’s own judgements ever being overridden. I shall discuss this sort of claim in more detail in the next chapter, in the section on privileged access. Here, let me just note that this is a dubious sort of claim for the physicalist; the paradigm which I am ultimately defending here holds qualia to be properties of states of the CNS, and tentatively hypothesises that they are identifiable with complex ‘physical’ properties. As such, qualia should in principle be capable of investigation by other parties; it should be possible to examine in detail the brain of an experiencer and ascertain whether in fact their beliefs about their own phenomenal states are accurate or not by seeing whether the appropriate set of phenomenal properties is actually being manifested or not.

So much, then, for the necessary incorrigibility of propositional awareness of qualia. What about contingent incorrigibility? It is a much more plausible thesis that, as a matter of fact, our beliefs about our own phenomenal states are virtually always true—that it would be a very strange set of circumstances indeed in which they were not. Even materialists, Armstrong says, 

... may even have some reason to think that introspective error about our current mental states will not be a conspicuous phenomenon. The causal chain between a mental phenomenon and the awareness of that mental phenomenon which it brings about is purely intracranial. As a result, it will not be subject to the interference brought about by unusual environmental conditions which can occur in sense-perception, especially a distance-sense as vision is. (Note that proprioceptive illusion is relatively rare.) (Armstrong and Malcolm 1984, 135)

However, that the causal chain is shorter and more secure from interference is surely not the most basic reason to think propositional apprehension of experience de facto incorrigible. At

---

26 What about all the great historical philosophers who seem to have held experience to be necessarily incorrigible? Often, I think, they should most charitably be taken to be thinking of phenomenal apprehension rather than propositional—and we have already agreed that this is, almost self-evidently, ‘incorrigible.’ Thus we have Hume: “Since all actions and sensations of the mind are known to us by consciousness they must necessarily appear in every particular what they are, and be what they appear” (1973, 190); and Descartes: “[C]ertainly it seems to me that I see, I hear, I feel heat. That cannot be false; that is what in me is properly called sensation” (1911, Vol. I, 153). The problem with this interpretation globally however, as Sellars pointed out (1956), is that it is hard to see how phenomenal apprehension could ever form the basis of foundationalist epistemological claims.
the root of this intuition, I feel sure, is the sense that we would ‘know’ if our cognitive attitudes were out of sync with our experience—that it would be very odd, even almost impossible, to attentively see a green field or feel burning pain, and believe that these experiences were absent or different. This is because the tokening of these phenomenal properties is itself a form of apprehension of those properties—to have an experience is exactly to feel a certain way. Thus, a “basic statement” about an experience which is false will typically be known to be false, since in some non-propositional way we already ‘know’ what the experience is like; for basic statements about qualia to be wrong would be for one’s own propositional awareness and phenomenal awareness to be in conflict. It would be analogous to reading a page oneself and simultaneously hearing it read out loud; any discrepancy between the two would be noticed immediately.

Further, continuing the analogy, one wants to say that any discrepancy would be settled in favour of experience—just as, presumably, looking at the words on the page would usually be considered more reliable than hearing them spoken … you might have thought you heard the reader say “apple pie” but the evidence of your own eyes shows you it must really have been “dappled sky,” or that the reader made a mistake. Just why we should consider phenomenal apprehension epistemologically favoured over propositional apprehension in this way is not entirely clear; but it is common—and surely accurate—to think of sensual experience as being far more “pressing” or “present” or “insistent” than mere propositional attitudes. Compare the experience of shooting pains in a tooth with the belief that one has toothache; it is certainly far harder not to pay attention to the former than to the latter.

In addition, beliefs have a ‘duty’ to be defeasible that experience does not have. Beliefs are either true or false, accurate or misleading—they have a normative ‘responsibility’ to change to match the way the world really is; experience, on the other hand, has no semantic properties, is

---

27 The claim here, note, is not that we could not have these experiences without forming a belief about them (evidence); it is that if we do form any belief about them at all, it will be an accurate one.

28 David Chalmers points to something very like this—though given extra bite by his anti-reductionism with respect to the phenomenal—and calls it “the paradox of phenomenal judgement” (1996a, ch. 5).
neither true nor false—it just is. It is reasonable to expect, therefore, that a successful mental system, such as our own, when it is functioning properly, will place the onus upon beliefs to change under the pressure of contrary apprehensions, rather than allow them to persist in the face of direct evidence of their falsity.

There is good reason to think, then, that propositional awareness of experience will typically be *de facto* incorrigible. Further, it is possible to make testable, and very plausible, predictions about how likely certain kinds of propositional awareness are to be faulty. For example, it seems plausible that *extra* beliefs about experience are more likely to be formed than false beliefs that directly contradict experience. That is, it is more likely someone might come to believe they see, say, the shape of a UFO behind certain lights in the sky than that a competent language user will insist those lights are red instead of green. Further, it is probably more likely that propositional mistakes will be made about less attended to qualia than about those most at the centre of attention; for example, it should be predictable that fewer false basic beliefs are held about severe pains or strong smells than about, say, motions in the corner of one’s eye, or faint sounds.

4. CONCLUSION

So, can we call our apprehension of qualia “certain”? If we restrict ourselves only to phenomenal awareness, then it is clear and unproblematic that this awareness is, necessarily, ‘indubitable,’ and ‘evident,’ and ‘incorrigible.’ On the other hand, we have found it highly unlikely that *any* subset of propositional apprehension of experience is necessarily certain. Some “basic” beliefs about experience are probably *de facto* indubitable and incorrigible, and a wider class of beliefs seems likely to be contingently evident. Once again, a table summarises our results:

<table>
<thead>
<tr>
<th></th>
<th>Phenomenal Apprehension</th>
<th>Propositional Apprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indubitable?</strong></td>
<td>Certainly yes.</td>
<td>Probably contingently yes, for some basic beliefs.</td>
</tr>
<tr>
<td><strong>Evident?</strong></td>
<td>Certainly yes.</td>
<td>Probably contingently yes, for some basic and non-</td>
</tr>
</tbody>
</table>
Incorrigible? | Certainly yes. 
--- | ---

It seems worth emphasising at this point that, despite what seem to be widespread extreme views of the “certainty” of qualia—or at least despite the attribution of such views to qualo- philers by their critics—the careful must recognise various avenues for ‘error’ with regards to qualia (some of which we have not even touched on here). Thus, as we have seen, one may experience a certain quale and not come to (propositionally) believe or know that one does. Similarly, there will usually be many propositions true of one’s qualia that one does not come to believe and alternatively, it is possible to hold certain beliefs—even “basic” beliefs—about one’s qualia that are not accurate. Further, even non-propositionally, one can experience qualia but not attend to those qualia—not ‘notice’ them in some way. And finally, the qualia themselves may be in some way inappropriate—in some way ‘wrong’: for example, one might feel pain in a phantom limb, or conversely fail to feel pain when it is appropriate, such as in the presence of damaging and normally painful physical stimuli.

The epistemology of qualia, then, is unique because of the phenomenal nature of qualia—our knowledge of qualia is more or less constrained and backed up by our phenomenal awareness of those qualia; nevertheless, it is not an arena of perfect certainty.
Chapter 6: The Privacy of Qualia Apprehension

Finally, before moving onto more overtly metaphysical matters, I want to consider the supposed privacy or subjectivity of qualia—the profound asymmetry that is often said to exist between first-person and third-person apprehension of experience. In the tradition of Descartes, this is often thought to be a metaphysical distinction, at least as much as it is an epistemological distinction between modes of apprehension. The fact of “first-person authority” is sometimes held to be the marker of a different kind of stuff than dull old objective matter; thus, for example, John Searle focuses upon the “subjective mode” as being “an ontological category, not an epistemological mode. … [T]he ontology of the mental is an irreducibly first-person ontology.” (1992, 94–95). Arguably, philosophers like Thomas Nagel (1974), Frank Jackson (1986) and David Chalmers (1996a) have also been members of this tendency. If only for the sake of focus, I shall avoid confronting this important kind of claim head-on; I hope, however, that my comments about what I take to be quasi-epistemological issues will make the “privacy” of qualia also seem less ontologically mysterious. I do recognise some residual metaphysical strangeness in this area, however, which I think is wrapped up in the very notion of a phenomenal property—this crops up especially in the notion of ‘perspectivalness,’ which will be discussed below, and I shall also return to the issue in my concluding chapter.¹

¹ I shall also, for a variety of reasons, cravenly ignore issues to do with the private language argument (see Wittgenstein 1958, 1980; Rhees, 1968; Budd 1989, Chapter 3; Kripke 1982). The argument itself (or set of arguments) is highly difficult to adequately interpret and assess and so I do not do so for reasons of space; furthermore, the question of the privacy of qualia is of course prior to that of the possibility of a ‘private language’ which denotes those qualia, and it is this prior question to which I devote my attention here. Finally, this question is perhaps less pressing for me than it might otherwise be for two reasons. First, a large part of Wittgenstein’s target for the private language argument is the epistemologically foundationalist element of the Myth of the Given (see Budd 1989, Ch. 3), and this is something I have already disowned. Second, as I have explained above, I do not hold that the names of external properties—such as redness—take their meaning derivatively from the names of phenomenal properties; for example, on my view, someone (such as a colour-blind person) could still correctly use the word “red,” despite the fact that they do not experience phenomenal redness. On the other hand, I do think that we sometimes want to name, or otherwise linguistically ‘point to,’ sensations—phenomenal properties and experiences—and that this is perfectly possible: this claim does, at least prima facie, come into conflict with Wittgenstein’s conclu-
Since privacy is clearly a concept that requires analysis in this context, I shall distinguish between four different uses of the term which I think have been most influential: privacy of possession, privacy of observation, privileged authority, and “perspectivalness.” None of these are, I shall argue, especially weird or surprising properties of qualia, given that qualia are phenomenal properties in the way I described in Chapter Three.

1. PRIVACY OF POSSESSION

One important notion of privacy is that something is private if it is somehow possessed by—or restricted to—only one individual (or a select few); this is just the general sense in which a house, a toothbrush, a vital organ, or a box at a sporting event can be private. (I say “general sense” because I do not mean particularly that only a special few have the right to use something private, or that the rest of us have a duty not to, or that some kind of force or threat of force is involved … just that for some reason or another access to or use of—the private object is highly restricted. This may perhaps be a somewhat metaphorical use of the notion of “possession,” but I think it is still straightforward enough for our purposes.) Since qualifiers are parts of my body (in virtue of being states of my CNS), and since all of my body is, after all, my body (and no one else’s) it is not at all surprising that qualia are private in this way. My qualia are private, in this sense, in a sense similar to that in which my liver is private: they are ‘possessed’ by me alone.

Furthermore, that quale-tokens—in virtue, merely, of being property tokens—cannot (normally) be shared by two individuals is, when one thinks about it, not in the least bit odd; in fact, it would be far more peculiar if it were not so. In this sense, as Seager puts it, “the privacy of qualia is no different than the privacy of any other property” (1991, 142): no more can two people experience the same quale-token than, say, two cats instance the same instantiation of the property of blackness.

Several of those who attribute the privacy of possession to qualia, however, or discuss such
Chapter 6: “The Privacy of Qualia Apprehension.”

attribution (Ayer 1959, Rorty 1965, Searle 1992, Unger 1990), have typically wished to make a rather more *prima facie* surprising claim—that (token) qualia are *essentially* privately possessed, in some way that things like toothbrushes and box seats are *not*. Thus H.H. Price (1950, 274) read “private” as meaning “knowable only to one mind” (“knowable,” note, not just “known”), and Ayer wrote that “we are obliged to deny that any sense-datum can be experienced by more than one person” (1940, 136). As John Searle put it succinctly, the view is that “leg transplants are possible; … pain transplants are not” (1992, 94).

Part of the feeling behind such notions may be a result of the historical blurring of the fact that qualia are *properties* rather than particulars (and experiences are collocations of property-tokens, rather than phenomenal individuals): thus I may be willing to share my toothbrush (or leg), but it would be much harder for me to share with or pass on to you my particular property instantiation of being, say, pinkish brown! Furthermore, Searle’s claim is, taken literally, probably *false* if the account of qualia I am defending is correct. It probably *is* possible in principle, in extreme non-normal cases, for more than one individual to undergo one and the same experience. Experiences, recall, are made up of qualia which in turn are properties of states of the CNS—these experiences can therefore be shared as long as (but only when) two individuals share one and the same token CNS state. Science fiction cases in which experiences might be shared, then, include conjoined twins who share sections of their nervous systems or potential non-human races in which different individuals somehow share the same, centrally located brain. Further, it is logically, and perhaps even empirically, possible to transplant pieces of the brain and in so doing to transplant experiences from individual to individual—thus two individuals might have the same temporally extended experience, but at different times (*A* will have the first half of the experience, *B* the second).

Clearly it is far more *difficult* to transfer ‘ownership’ of experiences (or, more exactly, qualifiers) than many other things, such as toothbrushes and limbs; but we should remember that it is still easier than the transfer of things like true memories and parentage, so qualia are not even in a class of their own in terms of degree of privacy of possession. In any event, there seems nothing especially puzzling or mysterious about this kind of qualia privacy.
2. PRIVACY OF OBSERVATION

This second notion of privacy (found in, among other places, Ayer 1959, Jackson 1977 and Dennett 1988) goes more to the heart of things: something is private in this sense—clearly one more or less expressly created for mental experience—if it can only be “directly observed” (by contrast with perceived, inferred, etc.) by one individual. Moore wrote that the “accepted view” of sense-data at the time was “that no sense-datum which any one person directly apprehends ever is directly apprehended by any other person” (1953, 43). More recently Searle has asserted that “there is no way that I can observe someone else’s consciousness as such; rather what I observe is him and his behaviour and the relations between him, the behaviour, the structure, and the environment” (Searle 1992, 97).

This doctrine of privacy, I think, embeds an important truth: that, as we have already discussed, qualia are phenomenal, and therefore are ‘detected’ in a different way than anything else of which we have knowledge—through a different ‘mode’ than, say, perception, inference or deduction. Further, it must presumably be the case that the experiencing of phenomenal properties requires their tokening—that is, only an individual in which qualia inhere can have phenomenal access to them.

Nevertheless, we should not be too liberal with the attribution of privacy of observation to qualia. Firstly, as we have seen, it is not necessarily true that only one individual can possess some token quale, and since to instance a quale is to phenomenally apprehend it then, pace Moore and Searle, there is some way in which two people could both “directly apprehend” (Moore) or “observe” (Searle) the same experience—though, admittedly, they in fact never do. Secondly, one would not want to make the claim (that Searle appears to be making) that our non-experiential knowledge of qualia can only ever be via some sort of inference from indirect pieces of evidence—that it is impossible to observe qualia from the third-person standpoint. At best, this claim illicitly presupposes the falsity of any kind of numerical identification between objectively observable features of the CNS and qualia; and from the perspective of a research

---

2 Or have there been actual cases of conjoined twins sharing relevant portions of the brain? I don’t know.
Chapter 6: “The Privacy of Qualia Apprehension.”

paradigm in which qualia are explicitly thought of as at least pseudo-physical properties of states of the CNS, it is flatly false.

There is, then, an asymmetry between the first-person and third-person modes of apprehension of qualia. This asymmetry renders first-person, experiential observation of qualia possible only for their possessor (or possessors); however, by itself it does not mean that qualia cannot be observed at all from the third-person perspective. This latter conclusion can only be derived by adding the as-yet unjustified (and certainly question-begging) assumption that qualia are non-identical with objectively observable properties.

3. Privileged Authority

James Cornman once distinguished six separate, philosophically relevant, senses of the word “private” (1971, 37ff.), but held that the only one really worth discussion was that “the person having an \( x \) is the final epistemological authority about whether the \( x \) exists (or whether it is the same as he thinks it is).” He notes, correctly, that to have “final epistemological authority” about something is different than having incorrigible knowledge: his example (39) is a baseball umpire, who has authority such that if she states that a batter is out, and doesn’t change her mind within an appropriate amount of time, then the batter is out no matter what evidence to the contrary someone else might have. But the umpire herself can overturn her own decision, as long as she does it within the allotted time. Cornman thus proffers the following definition:

\[
P \text{ has final epistemological authority about whether there are (are not) } x \text{s } \equiv_{df} \text{ If at any time } t, P \text{ believes that there is (is not) an } x \text{ at } t, P \text{ does not change this belief within an appropriate interval of time after } t, \text{ and } P \text{ understands what } x \text{s are at } t \text{ and during the}
\]

\[\]

3 The other five are: at most one person has each \( x \); each \( x \) is necessarily owned by one person; one and only one person can experience each \( x \); all \( x \)’s are experienced but none are perceivable through the senses; at most one person knows whether or not there is a certain \( x \).

4 “It is perhaps misleading to speak of this authority as ‘incorrugibility,’ for it is not true that a person may never be mistaken about whether or not he has something which is private in this sense. … Rather, this epistemological authority is based on his being in the best position to discover his mistake or to confirm his belief.” (Baier 1962, 98)
interval after \( t \), it follows that there is (is not) an \( x \) at \( t \). (40)

Once again, this is a notion of “privacy” that points out a real phenomenon when applied to qualia: though not perfect (as explored above), our knowledge of our own qualia is nevertheless ordinarily far better and more reliable than the judgements of third-party observers … since ordinarily, third-person judgements about qualia are based upon inferences from evidence only fairly loosely connected with sensation, such as verbal reports and other kinds of behaviour. A good actor (with a good make-up department) can give every external evidence of great pain, for example, without the corresponding qualia; conversely, extremely stoic individuals might be able to conceal any externally observable signs of intense pain.⁶

On the other hand, again, it is possible to press the thesis of privileged authority too far; in particular, it does not appear to be necessarily the case that the first person judgements of an experiencer must overrule those of other observers. Cornman’s definition is too strong for qualia; it is not the case, unlike with the umpire, that there is no prospect of overturning the judgement of the experiencer; and it is certainly not the case that the experiencer’s judgements are constitutive—that they make it the case that they are true, merely by being sincerely advocated.

First, as noted above, it is theoretically possible for more than one experiencer to possess the same quale, and these different experiencers might disagree in their judgements about the experience; one might think it a colour more close to blue than green, for example, and the other might not. Second, a completed “science of qualia” might ultimately have at least as much authority as first-person judgements. David Armstrong noted this in his 1968 *Materialist Theory of the Mind*:

Consider the case of a brain technician who has a perfect understanding of the corre-

---

⁵ Cornman also notes, in so many words, that the use of the notion of final epistemological authority renders a context referentially opaque: that is, *sahā veritate* substitutions of descriptions of \( x \) can affect the truth value of relevant sentences. See, for example, Quine 1980.

⁶ “It is obvious that until the brain-process theory is much improved and widely accepted there will be no criteria for saying ‘Smith has an experience of such-and-such a sort’ except Smith’s introspective reports. So we have adopted a rule of language that (normally) what Smith says goes.” (Smart 1959, 173–174)
lation between states of my brain and my mental states. Suppose, then, that I report “I seem to be seeing something green,” using the sentence as a phenomenological report on my visual experience. The brain technician is able to say from his knowledge of brain patterns that (a) I am not lying; (b) my brain is in the appropriate state for some other mental state; (c) there are disturbances in the brain-processes responsible for introspective awareness which would account for my mistake. On the evidence offered by the technician it ought to be concluded that I have made a mistake. (1968a, 109)

In short, if phenomenal properties could be accurately identified from the third-person using precisely calibrated scientific equipment—and there seems little reason at this stage to rule this out in principle (see Chapter Eight, section two)—and if that equipment could be shown to have at least us much accuracy as the propositional apprehensions of the normal experiencer, then conflicts between the two sorts of report need not be automatically settled in favour of the experiencer. The experiencer has, literally, “privileged access” to their own phenomenal states—a ‘way’ of ‘detecting’ them that others do not have—but this does not extend to absolute authority in their propositional judgements about those states.

4. PERSPECTIVALNESS

The fourth, and final, breed of privacy is a little harder to get to grips with. Mentioned in

7 I think, though, that to ultimately establish such a third-person science of qualia beyond a reasonable doubt one would need not only to map precise correlations between perceived and experienced properties but also to construct a satisfying theoretical explanation for why the third-person aspects of qualia go along with, or constitute, the subjective aspects. That is, one would want some reason to believe that the scientifically observed properties are in fact constitutive of, rather than merely correlated with, qualia. I discuss this point below.

8 It is also the case that, as Rorty put it, “inability to be mistaken does not entail inability to be over-ridden” (1965, 56)—that is, even if experiencers had ‘perfect knowledge’ of their own phenomenal states, there might still be conditions under which we would refuse to admit the sincerity of their reports, or come to believe that they are using language incorrectly. Rorty’s example is a case where someone undergoes painful stimuli, exhibits pain behaviour, suffers physical damage, but reports—between gasps and groans—that they feel no pain; in such a situation we would be most likely to judge that the reporter is lying, or does not know what “pain” means. (Rorty goes further and introduces Wittgenstein and the private language argument at this point, to suggest that the experiencer must actually not understand the meaning of the words she is using, since the criteria for their correct use are public and she is violating them.)
Frank Jackson 1977, 1982 and 1986, and in John Searle 1992, it is perhaps explored in most detail in Thomas Nagel’s *The View From Nowhere* (1986). Nagel calls the tension between “the perspective of a particular person inside the world …[and] an objective view of that same world … the most fundamental issue about … the relation of the mind to the physical world” (1986, 3). I take the following to be the main ideas making up the notion of “perspectivalness,” which is usually thought to apply exclusively to the phenomenal:

a) Mental experience possesses “perceptual aspects,” which can only be understood through the “senses” rather than via their “mathematical and formal properties” (Nagel 1986, 14).

b) Mental experience is such that it “can appear only to a particular point of view” (rather than to a “general rational consciousness”) (Nagel 1986, 15).

c) What is known perspectivally (subjectively) can only be apprehended in that way, and is not capable of being also apprehended objectively (Nagel 1986, 15–16). “Some things can only be understood from the inside” (Nagel 1986, 18). So no objective description of the world is a complete description of it, as Jackson argues (1982, 1986).

d) Nevertheless, “[e]ven though the concept of a mental event implies that it is something irreducibly subjective, the possibility remains that it is also something physical, because the concept doesn’t tell us everything about it” (Nagel 1986, 46). Indeed, Nagel tentatively considers himself a dual-aspect theorist, and Jackson explicitly holds qualia to be properties of brain states (Jackson 1986, 393).

Perhaps unfortunately, given the rather controversial position of the work of Nagel and Jackson within the contemporary debate, each of these four theses (can be interpreted such that it) seems fairly plausible to me. However I believe that, properly construed, the doctrine of perspectivalism is neither as wilfully mysterious nor as radically anti-physicalist as it has sometimes been thought (by, e.g., Paul Churchland 1985; Dennett 1988, 1991; and Shoemaker 1991 … or indeed, it seems to me, Nagel himself). I shall take the four claims one by one.

First, though I think it is rather confusing to refer to experience’s “perceptual aspects”—since qualia are *de facto* ‘known’ precisely through experience and *not* through their being perceived, and since by contrast objective, external properties like redness and squareness are per-
ceived—I take it that the contrast which is being alluded to in a) is that between phenomenal sensing and propositional knowledge. Thus, all of our objective, scientific knowledge of the world is propositional or mathematical—that the universe is expanding, that it is filled with blackbody radiation at a temperature of $2.7^\circ$ above absolute zero, that our planet’s atmosphere is mostly made up of nitrogen (78.09%) and oxygen (20.95%), and so on. On the other hand, the sensation of the temperature in a room, or of the colour of the wallpaper, is in itself not known propositionally—is not a proposition or mathematical formula—but is an experience.

Second, the claim that experience is restricted to a particular point of view (b) amounts to the following. As we have seen, phenomenal properties are felt only through being possessed; and typically only single individuals possess particular qualia—your visual sensation is de facto your visual sensation alone, and mine is mine. Further, an individual's sensory qualia are usually properties of states which make up the processing of information from that individual's sensory apparatus—a sensory apparatus which is located in some particular point in space-time and which has, thus, a relative perspective in no more mysterious a way than, say, the airport radar system at O'Hare has a different perspective on the world than that at JFK.⁹

What about the so-called Knowledge Argument (found in claim c) above): that perspectival knowledge can never be expressed non-perspectivally? The truth behind this kind of claim is simply that phenomenal apprehension is not the same thing as propositional apprehension—to know that a wall is red is, indeed, not the same as visually experiencing that redness; a colour-blind person could perfectly well do the former, but not the latter. On the other hand, it does not quite follow from this that qualia cannot be studied non-perspectivally—that their phenomenal natures cannot be described and understood objectively; that a complete description of the universe could not include a fully explanatory account of qualia. Qualia might be phenomenal because of their objectively observable properties, and nothing we know so far

---

⁹ The contrast with “general rational consciousness” is once again a reference to the distinction between phenomenal and propositional awareness: I suppose, analogously, that the radar system at JFK could 'know that' a certain aeroplane has taken off from O'Hare, though it would not be able to also ‘experience’ that fact, in the same way that O'Hare's system does.
gives us reason to rule this prospect out.

Consider, for example, the following (by necessity rough and partial) analogy: the relation between the property of transparency and the micro-structural properties of transparent objects. A complete description of the world at the micro-structural level, we might say, also captures the property of transparency since objects are transparent because of (i.e. in virtue of, rather than as a causal effect of) their quantum electro-dynamic properties (see, e.g., Feynman 1985, 107ff): that is, not only does the micro-structure of transparent materials fix their transparency, it also constitutes and explains it. Possibly, then, something similar is true of qualia—not necessarily that qualia are “higher-level” or structural properties, of course, but that a description of some other (objectively describable) set of properties will turn out in the end also to be a third-person description of qualia … and preferably that this identification will also explain why qualia are phenomenal in the way they are. That is, perhaps one day we will be able to look at the objective properties of some brain state and confidently say to ourselves, “A-ha, that must be a subjective experience of redness! How do we know? Well, because anything that has these kinds of properties must be phenomenal redness. Why’s that? Well, because ….” Of course, I have no idea how to fill in that ellipsis, nor do I even know with confidence that it can ever be filled in; all that is important at this point is that nothing in the notion of the perspectivalness of qualia need rule out the possibility.

Finally, what about claim d); the suggestion that the perspectivalness of qualia is compatible with their nevertheless being properties of physical objects? Naturally, we can wholeheartedly agree with this claim—but must we then be “dual aspect” theorists, which implicitly suggests that qualia themselves are non-physical properties? It seems to me that the main question here turns on what we take to define “non-physical” with respect to properties: I shall address this issue in Chapter Eight (but ultimately leave it an open question whether qualia are in fact non-physical). For now, let me just identify what seem to me to be two opposing considerations on this question. First, if the kind of identification of qualia with ‘physical’ properties that is countenanced above is indeed a live possibility, then it seems at least too early in the day to start insisting that qualia are non-physical. On the other hand, even then, we are already com-
mitted to a position that might usefully (if a little misleadingly) be called “dual aspect”—one in which qualia have both a phenomenal ‘face’ and an objective, third-person one.

5. CONCLUSION

In the end, then, is experience private? Certainly, in all the interesting senses of the term. However, in none of these senses is the privacy of qualia mysterious or difficult to accept (or at least, no more so than the very notion of phenomenal properties is to start with). In particular, in the first three senses qualia are not necessarily private, and in the last sense they are only necessarily phenomenal (which we knew already).

<table>
<thead>
<tr>
<th>Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Privately possessed?</strong></td>
</tr>
<tr>
<td><strong>Privately observed?</strong></td>
</tr>
<tr>
<td><strong>Privileged access?</strong></td>
</tr>
<tr>
<td><strong>Perspectival?</strong></td>
</tr>
</tbody>
</table>

This completes our consideration of the epistemology of qualia. The end result, in a sentence or two, is perhaps not all that surprising or original. Experience does indeed form a rather different epistemological ball-park than the rest of the world, and we have confirmed, upon consideration, that phenomenal apprehension is indeed in some sense “immediate,” that it is “certain” in the sense of being indubitable, evident and incorrigible, and that it is de facto highly “private.” Further, propositional knowledge of qualia can (rather simplistically) be said to be generally more “certain” than propositional knowledge of the “external world.” On the other hand, here is what I think might be considered ‘news’:

1) The epistemological asymmetries involved, and those epistemological aspects unique to qualia, are all basically the result of the phenomenal nature of qualia, and do not super-add any difficulties or implausibilities to this thesis (a thesis, which, I argued above, follows
Ch. 6 “The Privacy of Qualia Apprehension.”

3) Popular notions in this domain, like “givenness,” “immediacy,” “direct apprehension,” “certainty,” “privacy” and so on, require analysis—and I hope are clearer as a result of the work of these chapters—but in no cases need be attributing anything supernatural or implausible to human experience. In other words, merely the claim that qualia are immediate, certainly apprehended, and private, is (though unfortunately vague) not enough in itself to count as a reductio of the very notion of qualia.10

6. INTERLUDE: THE ‘MYTH OF THE GIVEN’

Before leaving behind my accounting of the epistemology of qualia, I wish to make a brief quasi-historical excursus to consider the influential and widespread notion that “the Given” is a mere myth, a discredited relic of the mistaken assumptions lying behind classical empiricism.

The attack on the notion of ‘sense-data,’ which I am concerned in this dissertation partly to reverse or temper, can be seen as an integral part of the twentieth century shift in Anglo-American analytic philosophy from the classic linguistic empiricism of Russell, Carnap and Ayer (in direct descent from the psychological empiricism of Locke, Hume and Berkeley) to what is now sometimes called “post-positivistic” analytic philosophy. This shift began in the early 1950s and was greatly influenced by a number of seminal works: in particular Quine’s “Two Dogmas of Empiricism” (1951), Wittgenstein's Philosophical Investigations (1954), Austin’s posthumous Sense and Sensibilia (1962), Ryle’s The Concept of Mind (1949), and Sellars’ “Empiri-

---

10 A salutary example is the way Rorty, in Philosophy and the Mirror of Nature (1979), builds an overly strong notion of incorrigibility into the very meaning of the term “mental” and then concludes that identity-theory must be false, leaving only dualism or eliminativism as possible options. (He went on to argue for the falsity—or at least linguistic emptiness—of dualism.) Dennett’s “Quining Qualia” (1988) is another example, considered in detail in Chapter Eleven.
cism and the Philosophy of Mind” (1956). As is well known these works, among others, apparently decisively dismantled many of the central preconceptions of logical positivism, such as the notion of “analytic truth,” the foundational epistemic role of sense-experience, the “logical atomism” of facts, the verifiability criterion of meaning, and so on. Of particular interest, for our present purposes, is the doubt they cast on the notion that ‘sense-data’ constitute a class of basic, immediately and certainly known facts to which individuals have specially privileged introspective access. This is what has become known as the “Myth of the Given.”

It has sometimes been thought that in refuting the epistemic notion that “something is immediately and incorrigibly presented to consciousness” it has been shown that therefore nothing is “immediately presented to consciousness”—that is, one might think, that qualia do not exist. This, however, is not the case; it is not a legitimate corollary of the post-positivistic rejection of the Given that we must also reject the very notion of “internal particulars” such as qualia and experiences. As at least Wilfrid Sellars recognised, one can consistently deny “the Given” (in the sense needed by Ayer) and yet still assert the existence of introspected, ‘immediate,’ mental particulars.\(^{11}\)

Sellars—the man who, after all, coined the phrase “the Myth of the Given”—held that there are two components to, or intuitions behind, sense-datum theory: the reification of ‘appearings,’ and the quest for epistemological bedrock. He was explicitly concerned to attack the second of these notions as the mythic Given, but held that there is at least one form in which the first kind of view—the existence of ‘appearings’ as particulars—is acceptable … and indeed, he ultimately concluded, necessary for a complete science. I, too, am defending a form of the reification thesis and not a version of epistemological foundationalism: in these fundamentals, then, my position is entirely compatible with the claim that “the Given” is mythical (though I have suggested above the various ways in which we might still want to talk of ‘the given’).

Sellars’ attack on the notion that ‘sense-data’ are epistemologically foundational is premised

\(^{11}\) Probably this has historically been obscured by the fact that Ryle, the later Wittgenstein, Quine et al. happen also to have been anti-realists, of various stripes, about the mental.
upon his very Wittgensteinian theory of “psychological nominalism.” He points out that, if ‘sense-data’ are to form an epistemic foundation then our apprehension of ‘sense-data’ must be a form of knowledge of certain facts: only then can it play a role in justifying our other beliefs. Thus, for sense-datum theorists (according to Sellars), it must be the case that sensing implies sensing as (e.g. sensing as being red), and hence if something is sensed as being x then the fact that it is x is non-inferentially known. Direct acquaintance with particulars, that is, implies knowledge of a certain sort of facts. “To say of a sense content—a color patch for example—that it was ‘known’ would be to say that some fact about it was non-inferentially known, e.g. that it was red” (1956, §4).

However Sellars, like the later Wittgenstein, identified the possession of a concept with the mastery of the use of a word. He held, therefore, that “all awareness of sorts, resemblances, facts, etc., in short all awareness of abstract entities—indeed, all awareness even of particulars—is a linguistic affair” (1956, §29). That is, one can only have knowledge that something is P if one already has the concept of P, which is to say, roughly, if one already has mastery of the word “P.” Hence, for Sellars, one cannot experience a red colour patch and then come to acquire knowledge of redness: one must first learn the use of the word “red” before one can even notice that the colour patch is red. Sense impressions by themselves, prior to language and the development of a conceptual scheme, cannot be knowledge, cannot even be conscious conceptual experience. Babies for example, according to Sellars, do not sense the fact that some patch is red … indeed, they do not sense any facts at all.

The essential point is that in characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says. (1956, §36)

---

12 As Rorty puts it, Sellars “may have been the first philosopher to insist that we see ‘mind’ as a sort of hypostatization of language” (1997, 7).

13 “[T]he idea that epistemic facts can be analyzed without remainder—even ‘in principle’—into non-epistemic facts, whether phenomenal or behavioral, public or private … is, I believe, a radical mistake—a mistake of a piece with the so-called ‘naturalistic fallacy’ in ethics” (1956, §5).
Knowledge—and in fact semantics generally, for Sellars—is importantly normative; particulars themselves (such as property-tokens or experiences), of course, are not.

Now, I am not at all sure that I wish to actually endorse the doctrine of psychological nominalism, but I am interested in showing that my position is reasonably comfortable with its results. And in fact I have already agreed, in Chapter Four, that we are aware of no facts at all simply in virtue of having qualia … or at least, that one can hold that position, if one wants to, consistently with everything I have to say about qualia in this dissertation. I agree, that is, that qualia are not “in the space of giving and asking for reasons” in the sense intended by Sellars; they are not pieces of knowledge upon which other knowledge claims can be constructed. Thus, as I say above, to phenomenally apprehend some quale is a non-semantic exercise distinct from the having of beliefs (or other cognitive states) about one’s own qualia. Qualia themselves are property-tokens, not beliefs; particulars, not facts. There is, as Sellars puts it, literally no such thing as a “veridical experience” (§1956, 7).

So the Given is a Myth, centrally, because qualia are not, in and of themselves, facts about qualia. However, many who attack the idea of the given seem to have thought that the central mistake embedded in this idea is exactly the idea that there are inner episodes, whether thoughts or so-called ‘immediate experiences,’ to which each of us has privileged access. I shall argue that this is just not so, and that the Myth of the Given can be dispelled without resorting to the crude verificationisms or operationalisms characteristic of the more dogmatic forms of recent empiricism. (Sellars 1956, §10)

Sellars himself more or less endorses the following story, which he attributes to the classical empiricists:

How does it happen that people can have the experience which they describe by saying ‘It is as though I were seeing a red and triangular physical object’ when either there is no physical object there at all, or, if there is, it is neither red nor triangular? The explanation, roughly, posits that in every case in which a person has an experience of this kind, whether veridical or not, he has what is called a ‘sensation’ or ‘impression’ ‘of a red triangle.’ The core idea is that the proximate cause of such a sensation is only for the most part brought about by the presence in the neighbourhood of the perceiver of a red and triangular physical object; and that, while a baby, say, can have the ‘sensation of a
red triangle’ without either seeing or seeming to see that the facing side of a physical object is red and triangular, there usually looks to adults, to be a physical object with a red and triangular facing surface, when they are caused to have a ‘sensation of a red triangle’; while without such a sensation, no such experience can be had. (1956, §7)

In a nutshell, for Sellars, ‘direct acquaintance’ with experiences is not sufficient for any form of knowledge (because the epistemic is irreducible to the non-epistemic) but may well be necessary for it (experiences—reified ‘appearings’—really do exist). Thus far, the Myth of the Given (at least in Sellars’ formulation) comes into no conflict with my own theory of qualia … indeed, it is even rather unexpectedly supportive of it.

There remains, however, one very important area of difference between myself and most of the well-known opponents of Givenness, including Sellars, and I shall conclude this section by drawing attention to it. Put in terms of my terminology, the problem with Givenness is that to experience qualia is not itself to propositionally apprehend those qualia, and it is propositional apprehension which supplies what are usually thought of as facts, which in turn might be candidates for epistemic bedrock. On the other hand, on my account, to experience qualia is to phenomenally apprehend those qualia, and, as explained above, this is surely of some epistemological significance. Sellars, by contrast, held that to even “notice” something requires that one have the concept of (i.e. mastery of the use of a word for) that thing: he was apparently quite ready to assert that animals and pre-linguistic human beings might have phenomenal experiences all the time but not notice them until they have acquired the words with which to describe them\(^4\): on my account, this is deeply wrong-headed. Pre-linguistic experiencers may be unable to conceptualise their experience (though see the discussion in Chapter Four, section three), but there is

\(^{14}\) Sellars explains this ‘noticing’ via his “myth of Jones” (1956, part XII onwards): Jones is the pioneering psychologist deep in misty prehistory who postulated inner thoughts and sensations as a way to explain human behaviour (just as molecules were hypothesised to explain the behaviour of gases); once we had the concepts we could then finally become aware that we have sensations, and so learn to non-inferentially report them. Experiences are theoretical entities which we learn to see directly: “What began as a language with a purely theoretical use has gained a reporting role” (§59). However, “the language of impressions was no more tailored to fit antecedent noticings of these entities than the language of molecules was tailored to fit antecedent noticings of molecules” (§62).
certainly a sense in which they notice it—a sense in which they are aware it is going on.\textsuperscript{15} Furthermore, it is natural and not obviously wrong to think of this phenomenal apprehension of our own qualia as some kind of \textit{constraint} or \textit{check} upon our propositional awareness: while it is not, as we have seen, impossible to make factual mistakes about our own occurrent qualia, the evidence from the tribunal of simultaneous phenomenal awareness of those very same qualia would tend to make this unlikely and quickly noticed.\textsuperscript{16}

Qualia then, in my view, are truth-\textit{makers}, not truth-\textit{staters}. They are in many ways ‘given,’ but they are not Given.

\textsuperscript{15} It was the failure to notice this form of apprehension which allowed those such as Ryle—and, today, Dennett—to deny that experiences even exist.

\textsuperscript{16} It seems that this kind of consideration could go some way towards resuscitating the foundationalist claims of old—perhaps even bringing back the Given in a different guise—but this is a big project which I shall not attempt here.
Chapter 7: Dimensions of Supervenience

We now turn our spotlight upon the metaphysics of qualia. The central question here is simply this: what sort of property are qualia? This in turn breaks down into at least two important issues:

1) What exactly is it to be a phenomenal property? How does ‘phenomenality’ work? I think this question can be treated as loosely analogous with questions like “what is transparency?” or “how does fluidity work?”

2) What is the relation between qualia and ‘the brain’? Are qualia ‘physical’? For example, are qualia (types or tokens) identical with complex physical properties of states of the CNS: is phenomenal redness, say, exactly the property of having charge-distribution $E$ and chemical properties $C$ (or vectors of those properties over some time period, or some kind of quantum gravitational field, or whatever)?

These are both extremely difficult questions and they are both best treated, at least at this stage of the research project, as empirical questions. Qualia are not concepts, relations or logical entities—they are properties of concrete physical objects; and the best way to find out about particular properties, prima facie, is to go out into the world and study them.¹ I shall not, therefore, attempt to answer these two questions here—to do so would be ‘Seven Planet Hegelian-ism’² of a particularly blatant kind (apart from being simply beyond my capacities). Instead,

¹ Having said this, it is possible that qualia will turn out just not to be open to empirical study. For example, qualia might (somehow) not be observable at all from a third-person viewpoint—might not be identical with any, simple or complex, empirically discoverable property of the brain. However, even in this extreme case, we shall have to look carefully at the brain in order to find this out (or at least, that is what I will argue below). If qualia were only apprehensible in any way from the first person, this would of course make them quite radically perspectival in the sense I outlined in the previous chapter … though still not quite as radically as Moore and Searle envisage since it still might be in principle possible to share token CNS states, and hence their properties, and hence the first-person apprehension of numerically identical qualia.

² This refers to the myth, popular among Anglo-American philosophers, that Hegel took himself to have demonstrated a priori that there must be seven planets in the solar system (when there are actually at least nine). It appears, however, that Hegel in fact argued for the more specific (and, as it happens, true) claim that there is no planet between Mars and Jupiter; further, he was arguing against those who held on a priori grounds that this (at the time) eighth planet must exist and doing so by showing there was an equally good a priori argument the other
rather as I did with the epistemology of qualia, I will try to clear some of the conceptual ground-brush out of the way.

Most of the philosophical attention devoted to the metaphysics of ‘qualia’ has historically centred broadly around the question of physicalism—that is, roughly, question 2) above. Relevant questions in this area include:

a) Are qualia physical or consistent with physicalism? (What does it mean to be “physical,” or for “physicalism” to be true?)

b) What is the ontological status of qualia? For example, are qualia additional to or somehow identifiable with already recognised properties of the CNS?

c) What is the causal status of qualia—do they have causal powers, and if so are these additional to or a subset of the already recognised causal powers of the brain?

d) In what way, if any, is qualia-talk reducible to the language of neuroscience?

e) In what way, if any, are claims about qualia predictable on the basis of neuroscientific claims?

f) In what way, if any, are claims about qualia explicable in terms of neuroscientific claims?

As should be clear, none of these questions (except the second half of a)) are unadulteratedly a priori; their answers will all presumably rely upon fairly detailed empirical knowledge about qualia—information which is simply not, at present, available. Instead of tackling these questions head on, therefore, what I will do for the next four chapters is try and make some progress on the conceptual underpinnings of these matters: I shall begin to develop a logical toolbox for discussing the relation between qualia (or experiences) and the brain. In the process, I hope to make clearer what is involved in adopting certain widely-held metaphysical positions about qualia, especially the view, which I believe is attractive to many, that qualia are:

i) not a counter-example to physicalism—there is nothing more in the head than the (physical) brain;

way. (See Beaumont 1954.) John Baker (e.g. 1996) makes reference to such “a manifested willingness to argue a priori on questions which are certainly empirical questions,” though he refers to it as “Ninth Planet Hegelianism.”
ii) not type-identical with physical properties of the brain;

iii) “supervenient upon” or at a “higher level than” the brain.

1. SUPERVENIENCE

I shall commence by discussing, in some detail, the supervenience relation. I do so for a variety of reasons—because supervenience has played an important role in some recent work in the philosophy of mind; because the supervenience relation provides a useful way to approach and characterise crucial contemporary debates about qualia inversion, zombies and the meaning of “physicalism”; because discussions of the supervenience relation have become so horrendously complex that some unifying conceptual analysis in this area seems particularly badly needed—but mostly because the family of supervenience relations is capable of providing a

---

3 The other direction I could usefully have taken at this point, it seems to me, would have been to go head-to-head on the question of identity theory. That is, I think there is a great deal of a priori conceptual work that can fruitfully be done on the possibility or otherwise that qualia may be type-identical with physical properties of the brain. My own inclination is to argue that such type-identities are not only possible but plausible (and the metaphysical work in this thesis is sometimes designed to make preliminary suggestions to that effect). However the battery of arguments against identity is formidable and, for many, persuasive, including, for example, Lewis 1980 (multiple realizability makes type-identities at least relative to populations), Putnam 1967 (mental type-properties—e.g. the property of occupying a certain causal role—are not strictly identical with the property of belonging to a certain physical type, since they are after all different properties), Davidson 1970 (mental types differ from physical types in being relative to background assumptions about meaning and rationality), Kripke 1980 (mind-brain identities are necessary or not at all, and they are not necessary) and Nagel 1974 ((some) mental states have a “point of view” that physical states cannot have). I have decided to focus upon the more theory-neutral notion of supervenience, and to postpone for the present work on identity theory, partly because I am aiming with this dissertation to establish a basic platform from which all work on qualia can be based—and an espousal of type-identities between qualia and “physical” brain properties strikes me as too controversial in some quarters to allow this—and partly because the supervenience relation is the more fundamental, subsuming the identity relation, so it seems reasonable to discuss it first.


5 “Supervenience’ seems over the years to have become an accordion-word: indefinitely stretchable, covering a bewildering variety of ideas related perhaps only by family-resemblance” (Post 1984, 163).
theory-neutral tool-kit for precisely describing the relation between qualia and the brain. That is, without committing oneself in these early days to strong claims about the reducibility or type-identity of qualia, it might still be possible to make progress with claims about the supervenience of qualia—to describe, with some precision, the way in which qualia plausibly covary with states or properties of the brain.6 Having got straight about supervenience I shall then (in Chapter Eight) apply this framework to my two metaphysical questions: how are qualia related to the brain, and how are they phenomenal.

The supervenience relation, intuitively, is simply this: S supervenes upon U iff fixing U fixes S. For example, some quale-type Q supervenes upon physical brain property P iff determining all the tokens of P fixes all the tokens of Q.7 There are other claims about supervenience, however, that are often made or assumed in the literature. I shall be interacting with, and often criticising, these dozen assumptions in what follows (the bracketed notes are shorthand for the conclusions I draw about these common beliefs):

- Supervenience is itself a special kind of dependence relation [False].
- The supervenience of S upon U is consistent with the irreducibility of S to U [True].
- Some supervenience thesis is an explanatory account of the mind-body relation [False].
- Supervenience relations are best categorised into the following three types: “weak,” “global,” and “strong” [Too simple as a taxonomy, too complex as an architectonic].

---

6 Here is what Jaegwon Kim once had to say on the subject: “We should note that most traditional theories seem to be committed to there being at least a supervenience relation between the mental and the physical. … [The supervenience relation] gives us a total asymmetric dependence of the mental upon the physical, and does this in a way that provides a realistic and reasonable account of the possibility of psychophysical reduction. … [I]t allows for certain restricted forms of psychophysical correlations and can make sense of them, without, however, demanding what is unrealistic, namely a pervasive system of biconditional psychophysical laws. Some will find the supervenience thesis incomplete as a mind-body theory just because it is consistent with a variety of incompatible mind-body theories, such as the psychophysical identity theory, the functional state identity theory, and even forms of emergentism. This I take as its strength; it is all we need, and it doesn’t make commitments where no commitments are called for.” (Kim 1982a, 144–146)

7 Or as Kim put it, mind-body supervenience is “roughly, the claim that the mental character of a thing is wholly determined by its physical nature” (Kim 1994, 575).
• Weak supervenience fails to provide a strong enough covariation relation for the defining of physicalism [True].
• Strong supervenience entails the reducibility of supervenients to subvenients [False].
• Global supervenience is the best prospect for non-reductive dependence, but may entail strong supervenience, and anyway is problematic as it permits the violation of (local) weak supervenience [False].
• Changes in a subvenience base are reflected in changes among supervenients [True, but somewhat complex].
• Changes in supervenients are reflected in changes in subvenents [Not always true].
• Each class of supervenients is in one particular supervenience relation with exactly one subvenience base [False].
• Particular supervenients (either local or global) are fixed by particular (local or global) subvenents [True, but somewhat complex].
• Supervenience relations themselves have no particular temporal duration [Sometimes false].

So, in this chapter I lay out a unifying conceptual analysis of the supervenience relation. It consists in a sort of three-dimensional matrix each of whose cells is a well-formed supervenience relation, and which is such that all of the core notions of supervenience in the literature can be assigned to one or some conjunction of the cells. This is not supposed to be a taxonomy of supervenience but an analysis of the underlying structure of the concept. The axes of the matrix are the specification of a covariation relation; the modal strength of that relation; and its spatio-temporal scope. It is my hope that this new model of supervenience can solve the problem of the undue proliferation of definitions of supervenience by delimiting and illuminating the ‘logical space’ available for such accounts, as well as by unifying, and making precise the relations between, the multifarious supervenience concepts already available.\(^8\) In the next

\(^8\) Several attempts at a ‘unified treatment’ of supervenience are already available, notably by Paul Teller (1984), Terence Horgan (1993), Jaegwon Kim (1984 and 1990), and Ansgar Beckermann (in 1992). However all such analyses that I am aware of, though often very useful, have centrally consisted in collecting together and taxono-
2. Covariance

A standard formulation of the intuitive core of supervenience is the following: “agreement in the subvening considerations (e.g. the physical) requires agreement in the supervening considerations (e.g. the mental)" (Teller 1995). We can begin by expressing this as the following simple determination relation:

\[ C1) \quad S_i \text{ (some supervenient) occurs whenever } U_i \text{ (some subvenient) does, but possibly also when it doesn’t).} \]

This, though of course it is not yet modally operated upon, captures the central notion that fixing the subvenience base fixes the relevant supervenients. That is, for some set of possible worlds still to be specified, it entails:

\[ C1') \quad \text{If two subvenients } U_i \text{ and } U_j \text{ are identical, then their respective supervenients } S_i \text{ and } S_j \text{ must likewise be identical, but not necessarily vice versa.} \]

In addition C1 has the widely popular property that it allows \( S_i \)'s occurrence without \( U_i \); other-

---

9 I am using the terms “supervenients” and “subvenients” in order to stay neutral on the topic of exactly what kind of thing the supervenience relation is a relation between. The options are numerous—supervenience and its kin have been taken, or could be taken, to be relations between at least the following (which are not supposed to be mutually incompatible):

a) Non-linguistic entities: properties, causal powers, objects or individuals, states of affairs, facts, events, patterns of similarity.

b) Linguistic entities: theories, predicates, descriptions, explanations.

c) Relational entities: functions, causal roles.

10 This notion is really the one at play in the literature where supervenience was first introduced (even before the actual term was coined), such as Sidgwick 1847, Moore 1922b and Hare 1952. Lewis’ dictum, “no difference of one sort without differences of another sort,” (1986, 14) and Davidson’s claim that “there cannot be two events alike in all physical respects but differing in some mental respects” (1970, 214) are classic modern statements of this as the central feature of supervenience.
wise, showing that the mental supervenes on the physical, for example, would also mean that fixing the mental fixes (at least a portion of) the physical.\textsuperscript{11} C1’s asymmetry is usually considered essential in order to take account of the phenomenon of multiple realizability: if some supervenient $S_i$ is “realizable” by a range of subvenients ($U_i$, $U_j$, $U_k$, …), then any covariation—it is commonly thought—must of necessity be asymmetrical in the form C1.\textsuperscript{12}

I take it then that C1 adequately expresses the central intuition behind supervenience; it is the most plausible construal of the covariation relation that lies behind—is the basis of—the supervenience relation. Nevertheless, there are two other common intuitive formulations of the supervenience relation that are \textit{not} entailed by any supervenience relation based upon C1, and it is instructive to consider the problems this raises. First, it is sometimes asserted that the supervenience of the members of class S on the members of class U means that fixing the state of all the members of U will fix the state of all the members of S: for example, it is often claimed that the supervenience of the mental on the physical means that “fixing the physical fixes the mental” \textit{simpliciter}. However from C1 it only follows that fixing the physical in a certain possible world—that is, specifying all the subvenients—necessitates the occurrence of \textit{at least} a certain set of mental supervenients in that world; it neither requires nor rules out the occurrence of \textit{extra} mental supervenients in that world, such as ghosts with beliefs or desires, additional epiphenomenal mental states in my mind, or perhaps even corpses that continue to have a mental life after death.

Second, supervenience is frequently identified with the claim that any change in some supervenient must be accompanied by a change in the subvenience base, or that any differences in

\textsuperscript{11} This would violate the principle of the ‘ontological primacy’ of the physical over the psychological—“That is, while it is possible for two things to agree completely in their psychological descriptions while differing in their physical descriptions, two things that agree completely in their physical descriptions could not disagree in their psychological descriptions” (Seager 1991, 99).

\textsuperscript{12} Though this claim is certainly true (as long as $U_i$, $U_j$, etc. are taken to be mutually incompatible), it is possible that it usually involves something of a misconception; in brief, typically when the subvenient of some “multiply realizable” $S_i$ is properly characterised, apparent cases of multiple realizability perhaps can be shown to be cases of \textit{symmetrical} covariation between some supervenient and a single equivalent subvenient.
supervenients are accompanied by differences in subvenients. This is not quite true. What follows from C1 is:

\[ C_{1'} \] Any variation in some supervenient \( S_i \) is accompanied by some change in the subvenient base away from \( U_i \), but not necessarily vice versa, if \( U_i \) is occurring.

However, if \( S_i \) is occurring without \( U_i \) then no change need occur. Thus, even if all mental states supervene upon the physical, it could still consistently occur that ghosts entertain a sequence of mental states without any relevant changes taking place in the physical. It should also be noted that even when some change in the base is required, no particular change is entailed. Hence, when \( S_i \) changes to \( S_j \), even if \( S_j \) covaries by C1 with \( U_i \), the base need not change from \( U_i \) to \( U_j \)—it could change to something completely different as long as the state to which it changes does not covary by C1 with any other supervenient.

So, for example, if some mental state—say, the thought that it is a pleasant sunny day (\( S_1 \))—covaries by C1 with some presently occurring physical state (\( U_1 \)), then any change in that mental state—perhaps into the thought that it is threatening rain (\( S_2 \))—entails some change in the physical, but as far as the C1 super-

---

13 “Such supervenience [of the mental on the physical] might be taken to mean that … an object cannot alter in some mental respect without altering in some physical respect” (Davidson 1970, 214). “Properties of type \( A \) are supervenient on properties of type \( B \) if and only if two objects cannot differ with respect to their \( A \)-properties without also differing with respect to their \( B \)-properties” (Horgan 1995, 778). “One set of properties is supervenient on a second set when they are so related that there could not be a difference in the first without there being a difference in the second, though there could be a difference in the second with no difference in the first” (Jones 1995, 860). “To say that considerations of one kind (e.g. the mental) supervene on those of another kind (e.g. the physical) is to say that there are, or can be, no differences in the first kind without there being differences in the second kind” (Teller 1995, 485).

14 The inference from C1 to C1’ relies upon the standard implicit assumption that the various relevant \( S \)-properties are mutually incompatible; for example, if \( U_i \) is sufficient for \( S_i \), then the occurrence of \( U_i \) rules out \( S_j \) (in the relevant spatiotemporal region) and vice versa. Suppose \( S \) occurs at \( t_1 \) but then changes to \( S_j \) at \( t_2 \); this means that, since \( U_i \) cannot be the case at \( t_2 \), then if \( U_i \) was occurring at \( t_1 \) some subvenient change must take place. But the C1 supervenience relation is quite compatible with \( U_i \) not occurring at \( t_1 \)—perhaps \( U_i \) does instead, or no subvenient at all—and in such circumstances no subvenient change need accompany that of the supervenients.

15 That is, when \( S \) changes to \( S_j \), \( U_i \) can change into any subvenient that does not necessitate some supervenient incompatible with \( S_j \), not necessarily to one that actually ‘fixes’ \( S_j \).
venience of any particular mental state goes, this change is absolutely unconstrained; it could, for example, consistently with the C1 covariation of both $S_1$ and $S_2$, be the utter destruction of the universe!

In a similar vein, under relation C1 a change in the base away from $U_i$ does not entail any change in the supervenient as long as the new state of the base does not covary by C1 with some other supervenient. So, to continue the last example, it is consistent with the C1 supervenience of the mental upon the physical that the universe could be utterly destroyed and we would continue to think that it is a pleasant sunny day.

This is not a happy situation. Many would think that some of the possibilities pointed to above—such as thinking corpses, and incongruous mental states accompanying the fiery demise of the galaxy—constitute counter-examples to supervenience. It might be thought that other components of the supervenience relation—such as proper modal quantification—would render C1 more acceptable (thus preventing problems with C1 from infecting the supervenience relations whose base it forms); I consider the modal dimension in more detail below, but here let me note that it is far from clear that it will be much help with this issue. No matter how extensive is the set of possible worlds in which some supervenience relation holds, it remains true that supervenients can occur without their respective subvenients in any of those worlds. Suppose the mental C1-supervenied on the physical by logical necessity, in all possible worlds: it is still consistent with this that in some possible worlds mental states exist epiphenomenally in the brains of the dead, or in universes of pure spirit; all that is required by C1 supervenience alone is that any two (regions of) possible worlds that are physical duplicates both contain at least those mental supervenients that are fixed by that physical base.

Two ways of ruling out thinking corpses by modifying the covariation relation at the base of the supervenience relation suggest themselves. Since a lot of the trouble is caused by the possibility of some supervenient occurring without its respective subvenient, the first and simplest method is to replace C1 as the basis of the supervenience relation with a symmetrical covariation relation instead, such as the following:

\[ C2) \quad S_i (\text{some supervenient}) \text{ occurs exactly when } U_i (\text{some subvenient}) \text{ does.} \]
Unlike C1, C2 does entail that supervenient variations are always accompanied by appropriate changes in the subvenient base, and that a complete specification of the subvenience base fixes a complete specification of the supervening level. However, since C2 means that supervenients must (within some specific set of possible worlds) be ‘coextensive’ with some particular subvenient, this solution apparently comes into conflict with the plausible principle of multiple realizability for higher level entities, as well as with the usual assumption that supervenience does not entail reducibility or inter-definability.

Probably a more acceptable move, and as far as I can tell the only alternative way out, would be:

a) to postulate a specification of the C1 relation for all the relevantly possible subvenients, and

b) to stipulate that the occurrence of some supervenient excludes the simultaneous occurrence of any other supervenient from the same domain in the relevant spatio-temporal region.

For example, suppose the mental supervenes on the physical, and that every relevantly possible physical state either determines the occurrence of some particular mental supervenient and no other, or determines the non-occurrence of any mental supervenient. Most of the standard

---

16 I mean to use this term rather loosely here, since I don’t wish to suggest that supervenience relations hold only between predicates.

17 This latter claim was an original assumption of Moore and Hare in introducing the concept, and was explicitly asserted by Davidson (1970, 214): “Dependence or supervenience of this kind does not entail reducibility through law or definition….” Kim (1990) makes it one of the three “desiderata” for supervenience.

18 Patricia Blanchette has pointed out to me that the wording of this clause is less than perfect for some cases. I am thinking of supervenient properties like colour or ethical value as canonical: if something is wholly red, it cannot also be wholly blue, and if some event is on balance morally bad it cannot also be morally good. However there are other examples that are less clear: could one and the same brain state determine more than one supervenient belief, for instance? I think that part of this unclarity is due to difficulties with the ontology of entities like beliefs as much as it is with the notion of domains of incompatible supervenients. More importantly, any vagueness in assumption b) is defused by its combination with assumption a): together these two principles mean that any given body of subvenients determines a certain set of supervenients and no other, even for blurry cases like beliefs.
claims about supervenience follow from this, while the possibility of multiple realizability is
held open. Presumably we would want to say that the physical state of corpses determines the
absence of any mental supervenient, and so supervenience rules out the dreaming dead. Keep-
ing constant any set of physical subvenients within a region fixes some particular set of mental
supervenients in that region, rather than just a subset of those possible. And it can be specified
that changes in the subvenience base must, in the relevant possible worlds, be accompanied by
appropriate changes in supervenients.

However it still does not quite follow that any change in the mental must bring about a
change in the physical: this is because C1 plus assumptions a) and b) (C1,ab) still allows some
supervenient to occur without the occurrence of any subvenient in the relevant spatio-
temporal region at all. Thus suppose S_i is occurring; this means that if some subvenient is oc-
curring at all, then it can only be one of those subvenients (U_i, U_j, U_k, …) upon which it su-
pervenes, since every other subvenient rules out the occurrence of S_i. Assume some subven-
ient is occurring, and assume that S_i changes to S_j. Since S_j too rules out every subvenient ex-
cept those upon which it supervenes, and since S_j cannot supervene on any state that deter-
mines S_i instead, then the subvenient state must also change. On the other hand, in a region
empty of relevant subvenients, supervenients can, on this relation, occur and change as much
as you like: in such “empty” regions we have still not ruled out, for example, the existence of
ghosts and spirits. Once again, tinkering with the other elements of the supervenience rela-
tion—modality and spatio-temporal scope—will not help us here, so if we are to remove this
unforeseen aspect of supervenience it must be through modifications of the covariation rela-
tion behind it.

It is not entirely clear that the supervenience relation requires the exclusion of ghosts, or—
which turns out to amount to the same thing—the absolute unconditionality of its formulation
in terms of supervenient change: indeed, it might plausibly be thought a virtue of this account
of the supervenience relation that the assertion of, say, the supervenience of mind on matter
does not, for example, entail the actual non-existence of God (assuming God would have men-
tal states). However a couple of moves would arrive at that end if desired (still without resort-
Chapter 7: “Dimensions of Supervenience.”

...ing to C2). One could make the additional stipulation that the absence of any subvenient determines the absence of any supervenient. Or one could specify that there are, as it happens, no “empty” spatio-temporal regions in the relevant set of possible worlds; for example, adopting what I shall characterise below as a “global” supervenience relation, in conjunction with assumptions a) and b), would probably achieve this effect. Either of these options can be captured by the following additional assumption:

c) Supervenients cannot occur in spatio-temporal regions empty of relevant subvenients (within the relevant set of possible worlds).

It is important to notice, however, that assumption c) (like C2) unavoidably makes supervenients de facto ‘coextensive’ with some particular finite set of subvenients (given a finite universe), which raises the prospect of the reduction of the supervenient domain to the subvenient—a prospect which is almost universally unpopular in many supervenient domains (such as the moral). I will return to this question below.

So this is the first dimension of supervenience: the basic covariation relation. I think it is moderately clear that virtually all supervenience relations extant in the literature fall along a single row in this respect: they are based upon the relation C1. However, this analysis has made clearer exactly what that covariation relation amounts to, and has made available two lesser variants that might turn out to be useful in some domains: C1_{abc} and C2.

3. MODAL STRENGTH

The second dimension of supervenience is the modal strength of the relation. Here there are four standard contenders (roughly construed):

M1) “Weak” necessity: that is, the covariance holds in the actual world but need not

---

19 It is ironic that global supervenience in particular turns out to play such a role in advancing this de facto reducibility, since that species of the relation has often been hailed as the one most likely to avoid reductionist commitments. See, for example, Haugeland 1982, Horgan 1982 or Chalmers 1996a.

20 In contexts where the two assumptions are not crucial, I shall simply call this relation ‘C1’ in what follows.
hold in any other possible world.\textsuperscript{21}

M2) Natural (or “physical” or “nomic”) necessity: that is, the covariance holds in the actual world and all naturally possible worlds—very roughly, all worlds where the fundamental, true laws of nature sufficiently resemble the actual laws.\textsuperscript{22}

M3) Metaphysical necessity: that is, the covariance holds in the actual world and all the metaphysically possible worlds—approximately, those where \textit{a posteriori} necessary truths (such as “water is H$_2$O”) still hold. This is a wider class of worlds than those mentioned in M2.

M4) Logical necessity: that is, the covariance holds in the actual world and all the logically possible worlds—roughly, those where \textit{a priori} necessary truths still hold. This, arguably, is a wider class of worlds than those mentioned in M3 and is the set of all possible worlds.\textsuperscript{23}

I believe that for us nothing much hangs on this particular way of picking out the important modalities, when another taxonomy could have been chosen. For example, some differentiate between natural and physical necessity (say, on the assumption that not all natural laws are physical—Chalmers 1996a is an example), while others deny that metaphysical and logical necessity describe different sets of possible worlds. I am most concerned with modality as an axis, rather than with which particular points along that axis are salient.

The consequences of combining some particular modality with covariance to generate an account of supervenience have already been fairly extensively worked out in the literature. For example Jaegwon Kim (1984, 57ff.) has argued that a modally weak supervenience relation does not provide anything that might be called a determination relation, since fixing the sub-

\textsuperscript{21} Jaegwon Kim distinguishes between “weak” and “strong” covariance in, among other places, 1990, 140ff.

\textsuperscript{22} David Lewis’s account of natural necessity (e.g. 1986) talks in terms of “similarity” between possible worlds—including their natural laws—but it is also possible to insist on identity of natural laws.

\textsuperscript{23} Some, such as Keith Campbell (1981) and Jaegwon Kim in his earlier days (1978) have insisted that supervenience is such that S\textsubscript{i} cannot follow deductively from U\textsubscript{i} (i.e. cannot be M4). Although there might be good reasons for this restriction on certain kinds of supervenience relation, or its use in certain domains, I do not think they are sufficient reasons to restrict the very definition of the relation.
venients does not fix the corresponding supervenients in even very similar possible worlds. I suggest that a weak C1 relation may actually count as one of determination (subject to spatio-temporal considerations discussed below), but agree that it is an extraordinarily thin-blooded one—perhaps, ironically, even more thin-blooded than Kim has (as far as I know) explicitly acknowledged. Consider a case where $S_i$ varies by C1 with $U_i$. Under modality M1 (weak necessity), knowing that $U_i$ has occurred or is the case does tell us that, as a matter of fact, $S_i$ has also occurred; it determines that $S_i$ is actually the case. However, it leaves us in the strange position that given even an infinitesimal change to our universe—say, the minor displacement of a hydrogen atom in deep space—$S_i$ might not have occurred, even if $U_i$ and indeed, say, our whole galaxy were utterly unchanged in that other possible world. Thus far Kim has noted; but in fact, there need be no physical change at all—a world with only a slight change to the mental, say, is still a different possible world than this one. Thus on an M1 account (even a spatially “regional” or “global” one), other possible worlds exist where the total subvenience base is identical to that in this world but the supervenients radically different.

Another capricious consequence of asymmetrical M1 covariance relations, which seems to go largely unnoticed, is that they are true whenever the antecedent is actually false. My mental states covary by C1 and M1 with, say, the mental states of unicorns—it is true that, in the actual world, if a unicorn has some thought then I have the same thought.

A covariance relation C1 modified under M2 (natural necessity) states something like the following: in any possible world where the laws of nature—ex hypothesi (though this claim is importantly indeterminate without a specification of the spatio-temporal scope of the relation: see the next section); yet all sorts of similar changes which did not actually happen, but which were possible—say, the very same change but happening a nanosecond earlier—could have utterly disrupted the covariance.

---

24 Another way of appreciating the strangeness of this situation is to reflect upon the fact that all the changes which actually do take place in the universe presumably have no effect on the covariation relation, ex hypothesi (though this claim is importantly indeterminate without a specification of the spatio-temporal scope of the relation: see the next section); yet all sorts of similar changes which did not actually happen, but which were possible—say, the very same change but happening a nanosecond earlier—could have utterly disrupted the covariance.

25 For non-global supervenience claims, this relies upon a tacit ceteris paribus clause, to do with antecedent conditions and so on: this clause could be tacked onto the covariance conditional, or alternatively the relevant condi-
covariance would be a change in the laws of nature themselves. This gives rise to the temptation to assert that the covariance relation operates because of the laws of nature—that the physical phenomena these laws describe somehow bring it about that $U_i$ is accompanied by $S_i$. Such an assertion is a plausible one in some domains, but it is not entailed by the use of the M2 modality. It could be that, for example, the laws of nature and, say, the supervenience of the mental on the physical have a common third cause, or some super-being could simply have brought it about by decree that such-and-such a covariance relation happens to hold in all physically similar worlds. Thirdly, obviously, the covariance relation could be a logically necessary one and thus true in a set of worlds of which the physically similar ones are a proper subset.

Logically necessary covariance (of either types M3 or M4, or perhaps of some other type again—I won’t worry about that here) carries with it a different kind of temptation: the temptation to ascribe the necessity of the covariance relation to some (quasi-)logical connection between the relata, such as that of identity, entailment, or some other mathematical relation. For example, we might say that the volume of a perfect sphere covaries by C1 (and C2) under M4 with the surface area of that sphere. More interestingly, Chalmers suggests that if physicalism is true then the totality of physical facts deductively entails the occurrence of all higher level properties, such as being alive or being lumpy, since these properties (ex hypothesi) are nothing over and above the physical facts (1996a, 41ff.). Whether this temptation—to read a logical connection into these supervenience relations—is to be resisted seems to depend upon one’s views upon the nature of logical necessity; for example, it may or may not be a counter-example to assert that some super-being may have simply decreed that $A$ and $B$ be correlated in all possible worlds and/or that we are just psychologically incapable of conceiving them apart.

We have now described a second dimension of supervenience, giving, so far, at least four different varieties of the relation (twelve if the covariation sub-variants are taken seriously).
Chapter 7: “Dimensions of Supervenience.”

Which modality is chosen in building a supervenience relation depends on the requirements of the domain in which it is to be used. For example, it might be that weak supervenience adequately describes the requirement for consistency in, say, ethical judgements (Hare 1952); the attribution of mental states under some instrumentalist account (Dennett 1987); or situations where theories are radically underdetermined by the data (Kim 1984). Davidson once suggested that weak covariation might be an appropriate way of understanding the relation between the truth of a sentence and its syntax (1970). On the other hand, it seems likely that most useful supervenience relations will make stronger cross-world claims, and be modified by M2, M3 or M4. Which is chosen might depend on whether the relation between subvenient and supervenient is thought to be, for example, a posteriori or a priori, and if empirical whether it is some kind of identity relation or merely a correlation. Thus, a priori relations (such as, arguably, those between moral properties and physical situations, or between higher-order and lower-order mathematical properties) will often be logically necessary; a posteriori identities (like that between water and H₂O, and between living things and certain kinds of self-organising and reproducing systems) might generally be thought of as metaphysically necessary; and empirical correlations (like those between certain brain lesions and changes in consciousness) could plausibly be dealt with under the rubric of (mere) natural necessity.

4. SPATIO-TEMPORAL SCOPE

A third axis for the supervenience matrix is necessary in order to properly specify some supervenience relations and to capture all the main varieties already extant: this is the dimension of spatio-temporal scope. Standardly the spatial scope contrast has been between, say, sets of properties of individuals (“local” supervenience) and of universes or possible worlds (“global” supervenience)—see, for example, Kim 1990, or Chalmers 1996a, 33ff.—though more recently notions like Kim’s “Structure-Restricted Correlation” (Kim 1989, 1992) or Horgan’s “Regional Supervenience” (Horgan 1982, 1993) have become more familiar. In my view the distinction between regional and individual supervenience is essentially unclear. The situation is best seen
as a spectrum of scopes lying between a single elementary particle and the totality of the universe: the terms “individual” and “region” operate within this spectrum more or less just as they do in ordinary language—that is, vaguely—and typically a region is larger and/or less structurally unified than an individual. However, for ease of exposition I shall simply talk about three varieties of spatial scope: individual/local, regional, and global.

Even now, the situation is not all that simple: there is no good reason why subvenient and supervenient should share the same spatial scope (though exponents tend to assume they do). There seem to be plenty of cases of modally quantified C1 covariation where the supervenient is larger than the base. For example, it seems plausible (though self-centred) to say that the total state of the universe now covaries in form C1 with natural necessity with a fact about me alone: the time that has elapsed since my birth (as measured in some quantum unit, say). Likewise, supervenients are often said to depend upon spatially greater subvenients: consider the property of “being a bank” which supervenes not just on the bank building itself, but also, roughly speaking, on the whole set of social institutions and conventions which make up banking; or wide content, which is said to depend not merely on my brain state but also on linguistic conventions and some facts about the world. All of this also suggests that even if both the subvenient and supervenient are either regions or individuals, they need not be the same region or individual—perhaps not even “overlapping” ones.

Similar considerations apply to the temporal scope of supervenience relations. The subven-
ience base can be smeared wider in time than the supervenient: “a person can change from married to single, or an item can change from your property to mine, at the stroke of midnight, when some prior legal decree takes effect—there being no physical change at the time, save, so to speak, the ticking of the clock” (Haugeland 1984, 6). In these cases physical changes back when the decree was promulgated must be included in the subvenience base. Alternatively, the supervenient could conceivably occur before the subvenient, or last beyond the cessation of the subvenient, or subvenient and supervenient need not overlap in time at all (as is standardly thought to be the case with causes and effects, which covary by C1 ceteris paribus).^{28}

The supervenience relation itself might also have some sort of temporal duration. Firstly, of course, there are cases where the antecedent of a covariance conditional (or each member of a biconditional) is true for only a confined stretch, or stretches, of time. Thus, we might loosely say, the actual supervenience of physical and mental states began at a certain moment in time (and will end at some other moment). More importantly, it could be that for some periods of time the covariation conditional is actually false, and at other times it is true. For covariance relations under modalities M3 and M4 this is presumably impossible (depending on one’s view of the basis for logical necessity); but if physicalism is false, this kind of temporal limit might obtain for some naturally necessary supervenience relations—if, for example, some nonphysical element were added to the universe at a certain point which, when combined with certain physical states, necessarily brings about some supervenience relation.^{29} Finally, weak covariance gives no reason why supervenience relations should not stop and start at any time: just as any possible change in the universe could end the covariation, so presumably could any actual change. In fact, an assertion of weak supervenience is importantly indeterminate unless it is given a temporal scope of this sort—that is, unless it is specified whether the covariation is supposed to hold for all time, for instance; or for all past time but an indeterminate period into

\[^{28}\text{Kim 1982b and Currie 1984 make related observations.}\]
\[^{29}\text{This is one interpretation of the position of the British Emergentists (described by Horgan in 1993). Another example of a time-restricted supervenience relation would possibly be that between the ‘green’ colour and, presumably, physical basis of a grue emerald.}\]
the future; for the next five minutes; or whatever.

This third axis—spatio-temporal scope—completes the supervenience matrix. Although we can now generate an indefinite number of somewhat different supervenience relations, for simplicity of discussion I shall assume that just three spatio-temporal scopes can do most of the work here:

S1) Individual/local supervenients and subvenients, temporally roughly coincident and covarying for an indefinite time into the past and future.

S2) Regional supervenients and subvenients, temporally roughly coincident and covarying for an indefinite time into the past and future.

S3) Global supervenients and subvenients, temporally roughly coincident and covarying for an indefinite time into the past and future.

This produces a total of at least twelve well-defined, core supervenience relations.\(^{30}\) However

\(^{30}\) Here, as a brief illustration of the way my model might be able to provide unity to the constellation of supervenience concepts available, I give recipes for some of the more interesting versions (as I, perhaps faultily, understand them):

a) “Davidsonian” Supervenience (Davidson 1970): C1, M1, S1.


c) “Weak Supervenience” (Kim 1984): C1, M1, ~M2–4, S1–3.

d) Haugeland’s “Weak Supervenience” (Haugeland 1982): C1, M2–4, S3.

e) “Strong Supervenience” (Kim 1984): C1, M2–4, S1–3.

f) “Global Supervenience” (Kim 1984): C1, M2–4, S3.

g) “Structure-Restricted Supervenience” (Kim 1992): C2, M2, S1.

h) “Local Supervenience” (Chalmers 1996a): C1, M1–4, S1.

i) “Regional Microphysical Supervenience” (Horgan 1982): C1, M2, S1–2.

j) “Mere Natural Supervenience” (Chalmers 1996a): C1, M2, ~M3–4, S1–3.

k) “Global Logical Supervenience” (Chalmers 1996a): C1, M3–4, S3.

l) “Physical Determination” (Hellman and Thompson 1975): C1, M2, S1–3.

m) “Bare Supervenience” (Blackburn 1985): C1, M1, S2.

n) “Non-Naturalistic” Supervenience (Moore 1922b, Hare 1952): C1, M1 or M4?, S1–2.

I shall not dwell on analysis of this list here, but it is interesting to note certain patterns, such as the close similarity between “strong” and “global” supervenience, for example (as predicted by Kim 1984), or the contrasts between Haugeland’s and Kim’s “weak” supervenience, or Kim’s and Horgan’s “structure-restricted” or “regional” supervenience.
so far, this purportedly exhaustive account of the supervenience relation has made no mention of supervenience as an important dependence relation.

5. DEPENDENCE

It is sometimes suggested that supervenience is intimately connected with an especially interesting kind of dependence relation; indeed, a major part of the philosophical attractiveness of supervenience has been the prospect of discovering a new kind of dependence, neither fishy reductionist nor foul dualist, hidden beneath the placid exterior of supervenience. But it has gradually become clear, in my view, that this project is doomed to failure, that supervenience does not contain or even shed light on the ‘appropriate’ sort of dependence.

Thus although, as we have seen, some of the relations that can be built from this conceptual tool-kit do provide a powerful determination relation, the kind of asymmetrical determination that follows from this is not of the right sort to count as “dependence” in this context. For example, none of the supervenience relations constructible on this model entail (though they are consistent with) a causal relation between subvenient and supervenient: a sphere’s volume supervenes upon its surface area under any modality, for example. Nor do any of them necessarily have an appropriate kind of metaphysical or explanatory asymmetry. Nor do any of them rule out some kind of causal or metaphysical joint-dependence on a third set of entities (where, for example, \(A\) and \(B\) covary, and both depend on \(C\) but not on each other). And, as

31 Paul Teller, for example, writes that “[f]or many of us the attraction of supervenience has been the prospect of having a way of saying that (for example) the mental is determined by, or boils down to, or comes to nothing more than the physical; while at the same time denying that the mental can be reduced to the physical in terms of any sort of explicit definitions or coextensive bridge laws” (1985, 93).

32 Kim, among others, has come to a similar conclusion. “…[I]t now seems to me a mistake, or at least misleading, to think of supervenience itself as a special and distinctive type of dependence relation…” (1990, 167). “[T]here is reason to think that ‘supervenient dependence’ does not signify a special type of dependence relation. This is evident when we reflect on the varieties of ways in which we could explain why the supervenience relation holds in a given case” (Kim 1994, 582).

33 Kim 1990, 142ff., makes similar points. See also Grimes 1988.
far as one can tell from what seems to me the quite vague usage of the notion of “dependence” in this literature, virtually any other kind of dependence is inadequate.

These problems with deriving dependence from supervenience can be treated as specific cases of the following observation: supervenience, it turns out, is everywhere, but supervenients are usually taken as depending only on some particular type of subvenient, specified in a particular way. Thus supervenience cannot itself be (a special kind of) dependence—to supervene is not necessarily to depend. For any given supervenient, it is possible to pick out several different subvenients: for example, my mental states covary, let us say, by a C1, M2 relation with the physical states of my region; by a C1, M1 relation with the mental states of a unicorn; by a C2, M3 relation at exactly time $t$ with my being alive; and so on. Further, for many of those postulated subvenient entities, the supervenient will covary with that entity in different ways under different specifications. For example, the psychological property of learning to avoid an electrified area of the floor supervenes upon the physical state of a rat in the following ways:

a) By $C1/M2/S1$ with some particular, completely specified, token process of a rat’s brain.

b) By $C1/M4/S1$ with some particular, completely specified description of the entire physical state of the rat between times $t_1$ and $t_n$ (during which time the rat learned to avoid the electrified patch, and exhibited this new behaviour pattern).

c) By $C2/M4/S1$ with a broad enough specification of the kind of physical behaviour which instantiates that sort of learning, along with any extra stipulations that might be needed to provide a functional definition of the supervenient property, such as that the organism in question be a self-guided system.

The task of deriving dependence from supervenience therefore involves picking out, not just the kinds of things which supervene, and their subvenients, but just which of the various supervenience relations involved are also dependence relations. It is not adequate to say that, for example, my mental states depend upon my brain—one must also specify which description of my brain is the relevant one, and thus which of several supervenience relations is at play.

It is still possible to argue that, although dependence is not derivable from the variously modi-
fied (modally and spatio-temporally) covariation relation I have described above, it is nevertheless an essential component of any proper specification of supervenience; that is, one might assert that dependence should be added to the three dimensions of supervenience to make a fourth. However I am not aware of any plausible, fleshed out candidates for ‘the’ relation of supervenient dependence which could be plugged in here. Kim has plausibly concluded that ‘the dependency relation’ is so metaphysically “deep” and “rich” that it cannot be captured by a fairly simple and generalisable relation at all (1993, 165 ff.). We are left, then, with the range of existing kinds of dependence: mereological dependence, between parts and wholes; causal dependence, between causes and effects; nomological dependence, captured by the basic laws of physics; meaning dependence, between linguistic entities; the identity relation, between something and itself; explanatory dependence, between an explanandum and its explanans; logical dependence; and so on.

In my view, dependency is better thought of, not as a component of the basic supervenience relation, but instead as a function of the domain in which the supervenience relation is used. Whether one thinks of a certain supervenient and some subvenient as having a relation of metaphysical dependency or not depends on one’s views about the metaphysical status of their respective domains, which in turn probably depends upon one’s position on, for example, realism, physicalism, or the unity of the sciences. These questions, it is plausible to say, go far beyond the issue of whether $S_i$ supervenes on $U_i$ or not—they are questions that remain to be answered, we might want to insist, even once the issue of supervenience has been settled.

We should note, however, that without some kind of dependence relation, supervenience relations are almost always unexplanatory. That is, supervenience itself merely picks out a certain kind of covariation or correlation—it does not explain why that covariation holds. In order to tell such an explanatory story, we usually need to pick out a certain kind of dependence—in particular, mereological, causal, or identity-type dependence.34

34 Kim (1990, 1993) suggests that the question of the supervenience of the mental on the physical is one about the phenomena of the mental, and is not intended to explain the mental. “Mind-body supervenience, therefore, does not state a solution to the mind-body problem; rather it states the problem itself” (1993, 167–168). Horgan
6. Conclusions

Here, in point form, are the most important conclusions about supervenience that I think are warranted by this discussion.

1) The apparently diffuse and manifold notion of supervenience actually has a relatively simple underlying logical structure which supports and explains the large number of uses of the concept. Which of the various possible sub-species of supervenience is to be applied on any particular occasion is a pragmatic matter, to be decided by features of the domain.

2) The class of relations of dependence is demonstrably a proper subset of the class of supervenience relations (on this model), and so supervenience cannot itself be a ‘new’ kind of dependence relation. It is possible that some ‘new’ kind of dependence could be added to the supervenience relation (as a fourth axis), but no such notion seems to be independently available. One could also argue that some existing sort of dependence (e.g. mereological or causal dependence) should be added to the very notion of supervenience, but I suggest that such an addition would probably be unwelcome.

3) It follows from this that the supervenience relation is descriptive rather than explanatory (but does not then follow that the notion is therefore of no use).

4) Exactly which supervenience relation is picked out as holding between some subvenient and some supervenient depends on their careful specification; it is inadequate (or at least very vague) to simply assert that, for example, “narrow mental states supervene upon the brain.”

5) The asymmetrical covariation of some supervenient domain with some subvenient do-

---

Stresses (1993, 577ff) that we need explanations of supervenience relations—that is, some account of why such a property is supervenient on a given configuration of physical properties. He argues that “putative supervenience relations that are themselves unexplainable and sui generis” cannot play a role in a broadly materialistic metaphysics, and advocates superdupervenience instead: “ontological supervenience that is robustly explainable in a materialistically explainable way.”
main must be supplemented with assumptions a) and b) above if one wants to ensure that particular supervenients (or their absence) are fixed by particular subvenients, and to make it the case that changes in the base are reliably associated with reasonable supervenient changes.

6) Assumption c) above is needed to rule out supervenients in regions empty of subvenients, and so to ensure that supervenient changes are always accompanied by subvenient changes; however, making this assumption brings much closer the prospect of the inter-definability of supervenient and subvenient, and hence perhaps the reducibility of the former to the latter.

7) Modally weak supervenience is, pace Kim, a determination relation, but it is perhaps even more thin-blooded than commonly thought and is importantly indeterminate without the specification of some spatio-temporal scope.

8) It is natural to assume that physically or naturally necessary supervenience holds “because of” or “in virtue of” the laws of nature, but this assumption is not always warranted.

9) Spatially global supervenience generally makes assumption c) true, and so, contrary to the apparent assumption of some in the literature, actually increases, rather than decreases, the prospect of the inter-definability of supervenients and some finite disjunctive subvenient.

10) The spatio-temporal scope of supervenience relations can be significantly more complex than is usually thought.

11) Supervenience relations that are merely weak or naturally necessary can themselves have temporal duration; they need not always be true.

I shall now return to the metaphysics of qualia: What sort of property are qualia? In particular,

\[\text{Roughly, the specification of a unique supervenient (within a domain) for all the relevantly possible subvenients.}\]

\[\text{That there are no supervenients in areas empty of relevant subvenients.}\]
I am interested in knowing whether they are ‘physical’ properties (and indeed, what this should 
best be taken to mean), and how it is that their phenomenality is to be explained. In the next 
chapter we shall progress some distance toward answering the first of these two questions.
Chapter 8: Supervenience and Qualia

Let us now apply to qualia the model of supervenience developed in the previous chapter. The aim is not so much to leap straight to an explanation of how qualia are instanced by the brain (a tall order!), as to try and specify exactly how (in what way) qualia are determined by the brain—or at least to cash out the consequences of qualia being related to the brain in certain ways. What we need to do now, then, is to describe the particular supervenience relation or relations that hold between the CNS and phenomenal mental properties. Once again, much of this is an empirically defeasible task: a definitive answer to this sort of question must in the end come, presumably, from work such as carefully controlled experiments tracking correlations between brain states and qualia and examining the dependence relations involved; or the construction of computer models or artificial brains and determining whether and how they produce qualia. Nevertheless, we shall try to make what progress we can a priori.

First, it is crucial to emphasise the relativity of the supervenience relation to different specifications of the base and supervenient. Blank claims like “qualia supervene upon the CNS” are, it turns out, almost uselessly vague, since they specify neither which aspects of the CNS are the relevant subvenients, nor in what way qualia supervene upon them. Thus, for example, nothing necessarily follows, from such an under-explicated claim, about, say, the relation between qualia and the brain in other possible worlds (on any particular accessibility relation), the truth of physicalism in the actual world, the status of the identity thesis, and so on. Let us therefore make a start on making the supervenience thesis about qualia more specific.

It appears to be almost universally accepted these days that, under some modality (if only that of weak necessity), the total set of qualia experienced by some subject A at time-slice t (if any) supervenes upon the completely specified state of A’s CNS and physical environment at

---

1 As we have seen, supervenience is primarily a descriptive rather than an explanatory relation.
2 Of course, such experiments will hopefully benefit from—if not rely upon—conceptual work such as that found here, dealing with the supervenience relation and the nature of experience.
that time; that is, exactly replicate A’s CNS and environment, quark by quark,3 and, at least contingently, you will fix the phenomenal properties occurrent for A at that time-slice. Call this the global token supervenience (GTS) thesis with respect to qualia. Further, it seems clear that the C1 covariation relation underlying GTS will need to be supplemented by assumptions a) and b): as we have seen, C1ab allows us, where C1 alone does not, to rule out the possibility of sensing corpses and qualia-laden rocks within some set of possible worlds.

Metaphysically, however, GTS is a very minimal claim. For example, it is consistent with most forms of dualism and epiphenomenalism (as long as it is accepted by dualists, as it usually is, that qualia are related in at least a vaguely ‘lawlike’ fashion, if only contingently and in the actual world, to states of the CNS). It is also, of course, compatible with physicalism, and even type-identity theory; therefore it under-determines any one of these metaphysical theses. However the GTS thesis is only a starting point: in this chapter I shall examine considerations which might drive us to add to or strengthen GTS in one way or another, turning it into a supervenience relation that has more specific metaphysical consequences.

1. THE SCOPE OF QUALIA SUPERVENIENCE

The matrix for supervenience relations outlined above has three elements: covariation, scope and modality. Since it is, in a way, the most straightforward of the three, I shall begin by discussing which spatio-temporal scope is the most plausible for the supervenience of qualia upon the CNS. The question is, in essence, how spatio-temporally local are the subvenients for particular quale-tokens (and quale-types)?

At the risk of sounding like a broken record, I should re-iterate that we are in a largely empirical domain here. However, I think I am on fairly safe ground in asserting, on the basis of

3 I want here to capture the notion of “exact duplication.” I gather that physicists consider the distinction between two token quarks to be a difference that makes no difference—that is, it makes no difference at all to any of the laws or descriptions of physics which quarks make up the individuals involved. So, if you were to take every quark in my body, remove it, and instantaneously replace it with some other quark (extracted from the centre of the Sun, say), then when you had finished you would essentially have made no change in me at all.
(what I know about) what is already empirically known, together with certain a priori considerations, that different sorts of mental properties and states supervene with a different spatio-temporal scope upon the physical. Roughly, the propositional attitudes appear to supervene upon a wide region, or even globally, whereas the phenomenal aspects of mental life probably supervene much more locally. One way of appreciating the wideness of the supervenience bases for the propositional attitudes is to consider the so-called Twin Earth cases in favour of “broad content” (Putnam 1975a, Burge 1979, Pettit and McDowell 1986, etc., but also Searle 1983, Fodor 1986): my desire for a glass of water supervenes not only upon the physical states and propensities of my body but also upon the chemical composition of that stuff which my language community has christened “water.” Further, even considering only the portion of the supervenience base that is the brain, it now seems empirically quite unlikely that a desire for a glass of water, or one’s unreasonable fear of cats, is located in a particular, specific neural location—that beliefs and desires correspond even neurally with specifiable, delimited bits of the brain, rather than being spread across a massively parallel, distributed, connectionist-style information processing network.\(^4\)

By contrast, my (very lay) assessment would be that there is quite good provisional empirical evidence for fairly marked localisation of phenomenal states (as opposed to information processing). That is, it seems plausible to suppose at this point, the qualifiers supporting particular qualia are fairly well localisable regions of the brain.\(^5\) However, it seems to me, the empirical

---

\(^4\) The need to attribute contentful cognitive states to others holistically has also sometimes been raised as problem for the localisation of belief states and desire states within the brain: the point, as far as I understand it, is that cognitive states can only be attributed en masse to a system as a whole, rather than individually to any particular part of that system.


\(^6\) William Seager also expresses the view that “qualia locally … supervene, propositional attitudes supervene globally” (1991, 114), and compares this with Fodor’s well-known distinction between narrow (locally supervenient) and broad (globally supervenient) content (Fodor 1981b, 1991). As Seager puts it, “to deny that two qualitatively identical brains would support some qualitatively identical and phenomenologically accessible psychological states seems flagrantly to violate physicalist intuitions” (1991, 134).
data relevant to this claim are rather difficult to interpret with confidence, partly because of a widespread lack of experimental differentiation between *phenomenal properties, representations, and cognitive processing*. For example, it is well known that particular neurones in cortical area V4 respond to different perceived colours (rather than merely, say, light wavelengths); while those in area MT (sometimes called V5) encode different perceptions of relative movement against a static background (see, for example, Crick 1994, Ch. 11). However, it does not strictly follow from these data that these clumps of neurones have the relevant *phenomenal* properties—that they themselves instantiate the visual experiences of moving, coloured objects. For example, they may possibly be part of the *pre*-phenomenally-conscious processing of visual information.\(^7\)

2. THE MODALITY OF QUALIA SUPERVENIENCE

Rather more interesting and complex, from our point of view, is the issue of the *modality* of the local supervenience relation between qualia and the CNS. Upon this hangs, via the choice of a covariation relation, such issues as the truth of physicalism for qualia and the prospect of epiphenomenalism, and I shall try to draw out these implications here.

Let us begin by asking what would follow if it were the case that qualia supervene upon the CNS merely *weakly* (i.e. by M1—possibly only in the actual world). Such an account of qualia would make extremely dilute the claim that states of the CNS *determine* qualia.\(^8\) As we discussed above, it would mean that there may be some possible world physically identical to this one in every way—one in which, in particular, all the human nervous systems are quark for quark similar to those in the actual world—where the distribution of qualia is radically different.

\(^7\) More radically, and rather implausibly, it is logically and, I take it, still empirically possible that qualia happen not to be properties of those parts of the brain that encode the current “best” cognitive representations, but of some other part of the CNS altogether.

\(^8\) As even William Seager, an advocate of weak supervenience, admits, it “is to say that the ultimately completely physical nature of psychological phenomena is a fact but not a very significant fact expressing anything like the ‘essence’ of the psychological” (1991, 167).
(since in virtue of having a different distribution of mental properties, it is a non-actual possible world; and since weak supervenience allows it as a possible world\(^9\)). Further, it might mean that the weak supervenience of qualia upon the CNS yesterday tells us nothing at all about the supervenience of qualia, even in the actual world, tomorrow. If we mean to say that qualia always have and always will supervene in the actual world upon the CNS, then we need to specify that as part of the hypothesised supervenience relation \(\ldots\) and given that the determination relation involved is so weak (as compared with, say, a physical law or some other regularity that supports a wider range of—indeed, any!—counterfactuals), it is a little hard to see how we might reasonably justify this stipulation. As we saw above, M1 supervenience entails that there is a vast range of possible worlds which evolve only minutely differently than this one, but in which the supervenience relation between qualia and the brain may break down or change.

The claim that the supervenience of qualia is merely modally weak, then, is unattractive. One way of expressing the discomfort we might feel with the implications described in the previous paragraph is to note that the mere M1 supervenience of qualia presumably renders the thesis of physicalism false, since it means that there may exist physically possible worlds, even worlds very similar indeed to the actual one, in which qualia do not exist, or are inverted, or are possessed by corpses, and so on. Let us pursue this question, and use it to consider the viability of M2—physically necessary—supervenience for qualia.

\(^{a)}\) Physicalism

Physicalism\(^{10}\), as a first approximation, says that nothing exists which is not physical or entirely dependent upon the physical for its existence. I shall not attempt to more accurately define physicalism here—I won’t try to decide what “the physical” denotes,\(^{11}\) nor shall I directly ad-

\(^{9}\) This is true even if, as is presumably so, there exist possible worlds in which the physical states of brains are the same as in the actual world and so are the distributions of qualia; that is, I am not suggesting that M1 supervenience means that qualia weakly supervene upon brain states only in the actual world.

\(^{10}\) I am here treating “physicalism” as being synonymous with “materialism.”

\(^{11}\) Poland 1994, especially chapter 3, is a useful entrée onto this important issue.
dress questions about the “location” of causal powers. I shall just assume that the brain and those of its properties explicitly dealt with by the neurological sciences—such as charge, mass, chemical composition, and so on—are uncontroversially physical; and that there is causal closure at the “level” of the physical. What we need to consider here is which, if any, doctrine of the supervenience of qualia upon the brain is consistent with physicalism. Since physicalism is, for many, a firm constraint upon theorising about the mind, this result will similarly constrict which putative supervenience relations are the most plausible or attractive. Even better, would be to find a supervenience relation which implies physicalism—that is, which rules out competing metaphysical positions, such as dualism—and I think we may be able to do that as well.

Let us begin by considering the prospect that qualia supervene by physical necessity (M2) upon the CNS, surely a plausible starting point for the physicalist. (After all, presumably, the thesis of physicalism need have nothing to say about possible worlds—such as the merely logically or metaphysically possible—in which the basic laws of nature do not hold; change the laws of physics, and all bets are off for the physicalist.) Adopting M2 would mean that in any possible world with sufficiently similar physical laws the determination of qualia by the brain would be preserved.

We have already decided that the supervenience of qualia upon the brain is most likely to be based upon at least covariation relation $C_{1_{ab}}$. (I shall continue to assume that symmetrical $C_{2}$ covariation is ruled out by the possibility of multiple realizability.) However we should remember that there is also the possibility of adding assumption c) to this relation—that, for some reason, supervenients cannot occur in spatio-temporal regions empty of relevant subvenients. This condition, recall, is required in order to rule out the occurrence of extra supervenients in regions empty of relevant subvenients. It is, however, more problematic with respect to physicalism, and turns out to be closely tied up with modal considerations.

Suppose we adopt an M2 modality and do not insist that something like assumption c) is the case: then, in any of the specified possible worlds (including the actual one), things like disem-

---

12 The notion of “levels” will, as it happens, be dealt with in the next chapter.
bodied minds may or may not exist. This means that the thesis of physicalism is simply false in all these possible worlds (since physicalism is false of possible worlds in which supervenients are not fixed by the physical). On the other hand, suppose c) is asserted for each of the relevant set of possible worlds (the physically possible ones picked out by M2). In that case, the thesis of physicalism entails that disembodied minds cannot exist in any possible world relevantly physically similar to this one: for example, the truth of physicalism proves that God not only does not exist but could not have existed in a universe like this one—there is no possible world in which the basic laws of physics hold and God or other spirits exist, even utterly epiphenomenally. In other words, physicalism (at least with respect to qualia) would be true in every physically possible world. This is surely too heavy a burden for physicalism—an a posteriori, empirical doctrine—to wish to bear alone. As David Lewis puts it:

Materialism is meant to be a contingent thesis, a merit of our world that not all other worlds share. Two worlds could indeed differ without differing physically, if at least one of them is a world where Materialism is false. For instance, our Materialistic world differs from a nonmaterialistic world that is physically just like ours but that also contains physically epiphenomenal spirits. … Our world might be … a world where spirits are absent but not outlawed. (1983a, 362–363)

The typical reaction to this problem is simply to restrict by fiat the supervenience relation to naturally possible worlds that lack extra or “alien” properties or entities—that is, entities which do not exist in the actual world and which might be problematic to physicalism. This is the resolution adopted by Lewis (1983a, 1994), and he is followed in this by most of those supervenience physicalists who explicitly recognise this conundrum. For example, Hellman and Thompson deal with the issue simply by deciding that “in the absence of positive arguments for extra entities, Occam’s razor (sound scientific procedure) will dictate commitment to the

---

13 A similar worry has been expressed by others trying to get straight about supervenience physicalism: for example, Horgan 1982 and Seager 1988.

14 That is, in effect, saying physicalism is true iff such and such a determination relation holds between the physical and non-physical and, as it happens, nothing iredeemably ‘non-physical’ (a.k.a. awkward for physicalism) exists. Chalmers calls this the addition of “a second-order ‘That’s all’ fact to the supervenience base in the definition of materialism” (1996a, 41).
However, as a result of the preceding analysis of supervenient covariation (Chapter Seven), I think we are now in a position to do somewhat better—that is to say, to attain a similar result entirely in the language of supervenience, without the need of any apparently *ad hoc* extra assumptions. I suggest the following solution: the appropriate relation between qualia and brain, consistent with physicalism, should be defined in terms of not one but *two* supervenience relations. The first relation is a physically necessary (M2), $C_{1ab}$ supervenience of qualia upon the CNS; the second is weakly necessary (M1) $C_{1abc}$ supervenience. In more detail, the physically necessary relation states that:

P1) In any possible world where the laws of nature sufficiently resemble the actual laws, $S_i$ (some quale) occurs whenever $U_i$ (some state of the CNS) does, and

a) For every relevantly possible $U_i$ there is some particular corresponding $S_i$.

b) The occurrence of some $S_i$ excludes the occurrence of any other quale in the relevant spatio-temporal region.

“Relevantly possible” in condition a) is best defined as follows: a physical subvenient is relevantly possible iff it is consistent with both the actual laws of physics and all the initial conditions that happen to have been the case in the actual universe. This is supposed to pick out every physical state that may occur in the history of the actual universe, but does not require us

---

15 William Seager, among others, outlines exactly two options for the physicalist supervenience of qualia (1991, 211ff.):

a) $S_i$ supervenes on $U_i$ in every physically possible world—i.e. the (physical) necessity operator is placed within the scope of the existential quantifier over physical subvenients.

b) $S_i$ supervenes on some $U_i$ in every physically possible world, but not necessarily the same one across all worlds—i.e. the existential quantifier is placed within the scope of the necessity operator.

However, he concludes, for reasons similar to those given above, that both these are options are flawed; he adopts the weaker relation, but apparently hopes for a third, intermediate option. My suggestion below skirts between these two problematic solutions and hopefully provides such an intermediate relation.

16 That is, two supervenience relations when constructed using my schema; these two could of course be considered together as a single complex relation, if the reader prefers.

17 But see my *caveat* on this kind of clause above (Chapter Seven, footnote 18).
to make a stipulation for all possible physical states.\textsuperscript{18} The second needed relation is the following:

P2) In the actual world, $S_i$ (some quale) occurs whenever\textsuperscript{19} $U_i$ (some state of the CNS) does, \textit{and}

a) For every relevantly possible $U_i$ there is some particular corresponding $S_i$.

b) The occurrence of some $S_i$ excludes the occurrence of any other quale in the relevant spatio-temporal region.

c) Qualia cannot occur in spatio-temporal regions empty of relevant subvenients.

The combination of these two supervenience relations has the following results. A complete physical specification of the actual world fixes all of the qualia: it determines that the qualia that actually occur do occur, and in the proper configuration, while ruling out the occurrence of any extra qualia (in disembodied minds, such as those of ghosts and gods). On the other hand, it does not implausibly assert that disembodied qualia are logically inconsistent with the laws of physics—there may be other physically possible worlds (different from the actual one) in which ghosts exist and physicalism is false.\textsuperscript{20}

While this definition of physicalism does not run afoul of the intuition that supervenients may be “realized” by a variety of subvenients, it does entail that, if only a finite number of subvenients can occur during the life of the universe, then as a matter of fact there is a finite, determinate set of subvenients for any given supervenient. But surely this should come as no surprise! The more important plank of the anti-reductionist thesis holds firm, for those who

\textsuperscript{18} One might also want to add the requirement, in condition a), that the supervenient fixed by each subvenient is ‘reasonable.’ That is, one might want to rule out, as part of the thesis of physicalism, the possibility that some physical state that may actually occur but has not yet done so is associated with some bizarrely inappropriate supervenient (such as that a proliferation of mammalian life on Earth will accompany the last few moments of the universe). However such stipulations are not required to make true the more general physicalist assertion that fixing the physical fixes the non-physical.

\textsuperscript{19} This relation is probably best understood as having an infinite temporal scope.

\textsuperscript{20} By the same token, on this account physicalism can of course be true in other possible worlds, viz. those in which assumption c) holds for that world.
wish to build upon it: for any disjunctive set of subvenients for some supervenient, there may be some naturally or logically possible world in which that disjunctive set is not exhaustive. This is the sense in which supervenients need not be identified with their particular subvenience bases, analogously with the way that, for example, “being a democracy” cannot be identified with a complete listing of all the currently extant democratic nations. (Of course, some or even all supervenient types might nevertheless be identical with certain subvenient types; the account of physicalism given here does not rule this out.)

I take it, then, that this account of the relation between qualia and the brain—of necessity, it seems, somewhat more complicated than is often thought—is consistent with the thesis of physicalism. Does it, however, entail physicalism, or is it also consistent with its negation? I suggest that there is a case to be made that P1 and P2 together probably do rule out the negation of physicalism … at least as long as the physicalist is allowed to be a non-reductionist.

Together, the P1/P2 supervenience relations imply that the distribution of qualia is precisely fixed by the status of the physical. What more can the physicalist want? Well, it might be objected, the fact that qualia are determined by the physical does not necessarily mean that they themselves are physical. For example, Seager writes at one point that “determination by the physical does not make a property a physical property” (1991, 102). However, taken literally, this is not a problem for physicalism—the thesis of physicalism is not typically meant to assert simply that all properties are physical,21 but rather that they are either physical or determined by the physical. After all, standardly the physicalist is concerned with the relation between physical properties and properties in some other domain (such as the economic, biological, moral or aesthetic), and that relation need not always be straightforward identity. Is the property of being a good person a “physical” property? Surely not. Shall we take goodness to be a counter-example to physicalism? Most do not do so—and they don’t do so, in part, because goodness is appropriately supervenient upon the physical; it is determined, in the right way, by the physical properties and states of the good person.

21 At least not since the days of Democritus and Lucretius … well, perhaps Hobbes as well.
Is there, then, some consideration that might make us want to adopt a dualism or emergentism about qualia that we do not adopt for goodness or beauty? Many of the standard encouragements to dualism, it’s important to notice, are ruled out by P1/P2 supervenience; for example, this kind of supervenience means that qualia cannot be inverted in other possible worlds relative to this one, or absolutely absent. If some state of my CNS $U_1$ has a certain phenomenal property, then in any physically possible world it has that property (though it remains logically possible that qualia might be inverted or absent—in worlds with similarly arranged ‘atoms’ as this one, for example, but with different or absent laws of physics). Thus, if physicalism is true of the actual world, actual physical states have the phenomenal properties they do by (at least) physical necessity, just like all their other properties: qualia, on this account, are no more loosely determined by or connected to the physical than any other kind of physicalist property.

The only incentive for dualism that I can see then, given P1 and P2, is the antecedent assumption that, for some reason, qualia are not physical, and that they are ‘not physical’ in some stronger sense than that in which goodness is not physical. For example, that qualia are phenomenal, and thus in a sub-class of properties by themselves in this important respect, might be taken to ground an important duality, perhaps akin to Descartes’ influential dichotomy between the thinking and the extended. I, however, cannot see how to make this view much more than ad hoc. Why, for example, should phenomenality be considered the hallmark of the “non-physical,” although it were determined by the physical, when, say, normativity or intentionality or quantum indeterminacy—on the face of it, surely, equally or more inimical to our ‘folk physicalist’ intuitions—are not?

I conclude, then, that if qualia supervenied upon the brain by relations P1 and P2 then physicalism would be true with respect to qualia. We must now face the following question: is the physicalist supervenience relation between qualia and the brain, P1/P2, actually true of qualia?

b) Zombies, Qualia-Inverts, and Epiphenomenalism

Though it would be nice if P1/P2 were true, since we have seen that this complex relation is
both required by and implies physicalism, there is reason to think it might not be. Empirical work has a bearing here, in a way I shall explore in a moment: briefly, if it can be shown that, for example, physical properties constitute qualia, and that this constitution is implied by the laws of physics, this would tend to show physicalist supervenience must be the case.\textsuperscript{22} However, there is also an important \textit{a priori} problem here which we shall deal with first. In essence, can we show \textit{a priori} that the P1 supervenience of qualia does not hold for the actual world? Because if so, obviously, then we must already know that qualia physicalism must be false: we must already know that qualia supervene upon the physical in a way which is inconsistent with the truth of physicalism.

The question to ask here is the following: if actual brain state $U_1$ determines actual quale $S_1$ (say, a certain shade of phenomenal green), is there a physically possible world in which $U_1$ corresponds with some other quale (such as phenomenal magenta)? If so, of course, the determination involved cannot hold by physical necessity and so cannot be an instance of P1 supervenience (and so physicalism must be false). This question is often debated through discussion of the possibility of inverted or absent qualia; in particular, I shall focus upon those theoretical creatures physically and behaviourally totally identical with human beings but which completely lack qualia, and which are often called “zombies.”\textsuperscript{23}

So, are zombies and qualia-inverts physically possible? It is sometimes suggested that if inverted or absent qualia are even possible (simpliciter), then physicalism is false\textsuperscript{24}; it should be

\begin{itemize}
\item [\textsuperscript{22}] Empirical work might also, of course, impact upon the claim that qualia actually are actually correlated with brain states.
\item [\textsuperscript{23}] Robert Kirk was probably one of the first philosophers to make much use of the notion of zombies to criticise qualia physicalism (1974 … though now—in \textit{Raw Feeling}, 1994—he defends physicalism against zombies), and Dennett in particular has always been a staunch ‘zombiephobe’ (1982, 1991, 1995a). Dennett has called the attention given to zombies by philosophers “an embarrassment to our discipline” (1995b, 325). David Chalmers discusses zombies at some length in \textit{The Conscious Mind} (1996a, especially 94–99 and Ch. 4). A symposium on “zombie earth” was published in the \textit{Journal of Consciousness Studies} 2:4 (1995), in response to Todd Moody’s keynote article “Conversations with Zombies” (1994).
\item [\textsuperscript{24}] “Physicalism is itself a modal doctrine, expressible as something like: any two physically possible worlds that are physically indistinguishable are total duplicates …. But …the merest possibility of epiphenomenalism confutes
\end{itemize}
clear that this is not quite right. It is logically possible that pigs can fly, borne up by their invisible immaterial pet angels; nothing follows for pig physicalism in the actual world. What is relevant is whether such phenomena are physically possible. Thus, that we can imagine zombies or qualia-inverts is not enough—certainly the physicalist can admit that there are possible worlds in which such things go on. Rather, what is of interest is whether, in a possible world very like this world, with all of physics held constant, zombies or spectrum inversion could (logically) possibly occur. What can we say, for and against, about this possibility?

David Chalmers is arguably the paradigmatic advocate for zombies, and his position is based on the following reasoning (1996a):

i) Given the laws of physics, for any structural or functional property, a complete description of any physical instantiation of that property makes it inconceivable that it should not be an instantiation of that higher-level property.

ii) Virtually all non-physical properties are structural or functional; that is why they supervene upon the physical by physical necessity—they are fixed by the physical in all physically possible worlds.

iii) Qualia, by contrast, are neither structural nor functional properties, and a complete description of a physical instantiation (or correlate) of a phenomenal property, even given the rules of physics, would not rule out the conceptual possibility of that phenomenal property being absent.

iv) Therefore, the supervenience of consciousness is weaker than physical necessity (and so weaker than P1).

This argument (and other less sophisticated versions of it) is mockingly labelled in some this doctrine” (Seager 1991, 167).

25 Thus David Chalmers (1996a, 33, 86–87), for example, prefers to define physicalism in terms of the logical entailment of supervenients by the physical, but here by “the physical” he means explicitly to include the laws of physics. That is, qualia would be logically entailed given that this is a physically possible world (but, presumably, not otherwise). However he then quickly segues, as far as I can see illicitly, to the claim that “the question is whether the notion of a zombie is conceptually coherent. The mere intelligibility of the notion is enough to establish the conclusion” (1996a, 96).
quarters the “argument from the failure of imagination.”

Though cruel, this attack, I think, does find its mark: the argument does seem to boil down to the claim (in premise iii) that, since we cannot imagine how qualia might be necessitated by the laws of physics, they are not so necessitated. Assuming (as seems plausible) that Chalmers is correct in asserting that structural and functional properties are entailed by their supervenience bases in all physically possible worlds, it still remains to be shown that only such properties are so entailed (and, even then, that qualia in fact will turn out not to be structural or functional properties of some kind).

What is at issue, in fact, is exactly the kind of dependence relation that holds between qualia and the brain. If it is a variety of dependence that necessarily follows from (or holds given) the initial conditions and the laws of physics, then zombies and qualia-inverts are impossible in any physically possible world, and so physicalism is true. Otherwise, physicalism is false. Thus, for example, suppose that qualia are caused by states of the CNS (in some fairly ordinary sense such that the effect is physically necessitated by the cause), or that they are identical with certain complex physical properties, or that they are constituted by a set of basic physical properties in a way analogous to transparency, acidity or liquidity: in any of these cases, zombies and qualia-inverts would be physically impossible. The ‘zombiephile’ must deny that any of these circumstances are the case, and it is not at all clear she is justified in doing so; surely it remains to be seen exactly in what way qualia depend upon the brain. The zombiephile, in other words, cannot just assume that the relation between qualia and brain states is ‘non-physical’; evidence for this hypothesis

---

26 This is, for example, how Patricia Churchland characterised it during her presentation at the conference “Toward A Science of Consciousness II” in Tucson in 1996, and I gather that the phrase was then already in fairly common currency in certain circles.

27 “… [A]s we have seen, if such doctrines as epiphenomenalism are just logically possible (and it is very hard to deny that they are) then there are physically possible worlds that agree on the physical state but disagree on the states of consciousness instantiated in them. These worlds are produced simply by breaking the non-physical causal relation linking physical states to states of consciousness. Since the link is not a physical feature of the world, altering it does not affect its status as physically possible” (Seager 1991, 213).

28 In the same way, we do not consider it physically possible that moving electric charges might not bring about magnetic fields, or that the actual referent of “the Evening Star” might not be the same thing as the actual referent of “the Morning Star,” or that H$_2$C$_2$H$_2$O$_2$ (vinegar) might not be acidic.
must be independent of, and prior to, evidence for the physical possibility of zombies. Zombies, then, are not evidence for the falsity of physicalism; they are merely a consequence of the (putative) falsity of physicalism.

It is sometimes argued that because one can imagine zombies or qualia-inverts in worlds physically identical with this one such things must be possible; therefore, on a priori grounds, qualia cannot depend upon the physical in any of the above ways (but must be “merely correlated” with states of the CNS). If we felt we really understood phenomenal properties this argument might have some weight; after all, Chalmers and others are quite right to insist that we cannot honestly imagine well-understood structural or functional properties like being a chair or being solid varying independently of the physical, whereas we find it easy to envisage inverted or absent qualia in a brain otherwise just like ours. However, in a case like this where we are just beginning to investigate the science of qualia, the argument from imaginability has no force. The fact that, for example, I can imagine the laws of physics remaining true and the universe at some point in the distant future beginning to collapse back in on itself has no impact one way or another on whether this is physically possible, or whether the universe must expand for ever. We don’t yet know whether either or both of these things are possible, and to find out we will have to go out and investigate the universe itself (such as by calculating its mass, calculating its laws more finely, or even discovering relevant new sources of force), rather than speculate a priori.

It might be objected that in the case of qualia—as distinct from the future trajectory of the universe—one can envisage both knowing everything there is to know about the physical universe and still imagining physically possible worlds containing zombies. However, this begs the question. If, for example, qualia were constituted by the physical then anyone who really understood all of the physical universe could no more imagine a physically possible displacement of

---

29 This would of course be a modus tollens from the conclusion reached in the previous paragraph.
30 Recent evidence has arisen suggesting it may be the case that absolute vacuum can have energy (because of its froth of constantly arriving and disappearing “virtual particles”), and that this energy might be sufficient to counteract the gravitational force which acts as a centripetal brake on the expansion of the universe.
Chapter 8: “Supervenience and Qualia.”

 qualia than she could a displacement of acidity or inflation rates. That we think we can imagine it is of no more relevance than the fact that many of our ancestors probably thought they could imagine the physical universe exactly as it is but where lead can be turned into gold through chemical action, or with a different distribution of infectious disease. The fact that we can imagine physically possible zombies, then, is epistemologically significant: it shows that we do not yet have a complete theory of qualia. It is not, however, metaphysically significant: it does nothing whatsoever to show that physicalism must be false with respect to qualia.

I conclude, therefore, that we cannot demonstrate from first principles that the zombie intuition is true, and physicalism false. Can we, on the other hand, show a priori that zombies must be impossible? This again would be a very powerful result, since it would demonstrate that physicalism must be true with respect to qualia. One way of doing this—indeed, the only way that seems immediately plausible\(^{31}\)—would be to show that epiphenomenalism is impossible: since we must accept that the possibility of zombies or qualia-inverts entails the epiphenomenalism of qualia,\(^{32}\) this would show zombies to be impossible by modus tollens.

The appeal to epiphenomenalism, however, is inadequate as an a priori reductio. There might be two reasons for thinking that epiphenomenal consciousness borders on the impossible or absurd. First, one might think that epiphenomenal properties are in some way alien to the natural world—that we would, so to speak, have to introduce the very notion just in order to apply it to phenomenal consciousness. But this is not the case … at least not in the most plausible construal of “epiphenomenal” in this context. We need not—indeed, should not—use

\(^{31}\) The only other major tack of which I am aware is that exemplified by Daniel Dennett (1991, 1995a, 1995b) who insists that creatures are (phenomenally) conscious just in virtue of being a zombie (or, in Dennett’s terms, a “zimboe”). This kind of story, however, is only plausible when combined with Dennett’s eliminativism towards qualia: it is therefore inconsistent with the results we have already established in, especially, Chapter Three. I consider Dennett’s eliminativism in more detail in Chapter Eleven.

\(^{32}\) If it were true that a physical replica of me would necessarily exhibit exactly the same behaviour as I do but could possibly lack any phenomenal conscious states, then this seems to show that phenomenal consciousness does not contribute to the causal chains that bring about behaviour—or at least that if they do, then those behavioural effects are causally over-determined.
the term to mean that qualia are not causally affected by the physical; it only means that phenomenal properties do not impinge causally back upon the physical. Nor need we yet insist that experiences have no causal powers at all; perhaps they may have causal effects upon each other. So really, what we are talking about here—what we mean by “epiphenomenal”—is just relative causal inertness: being causally inert with respect to $x$. And this kind of second-order property—relative causal inertness—can be found all over the place. The colour or opacity of billiard balls has no effect on the causal interactions between those balls and with the table (among other things); the texture of foodstuffs has no effect on their nutritional value (among other things)—broccoli is just as good for you whether puréed or not; tables have the same ability to support objects on their surface no matter what rigid material they are made from or how many legs they have (though these things are causally relevant in other respects); and so on and so forth.

An objection one might make at this point is that the putative epiphenomenality of consciousness is actually not, as I have presented it here, just different (if at all) in degree rather than kind from the relative causal inertness of all sorts of household properties. In fact, this objection runs, consciousness would differ from any other property yet discovered in having no causal effects at all (if the zombie intuition stands). If one accepts that states of phenomenal consciousness in the actual world supervene (by C1$_{ab}$) upon physical states of the CNS, then particular changes to the CNS will de facto mean particular changes in consciousness. If one further accepts that, in some sense, CNS changes ‘cause’ phenomenal changes then this generates a dilemma: either phenomenal conscious states have no causal effects either on the physical or on each other, in which case it is hard to see how they can have causal effects at all; or changes in phenomenal consciousness are causally over-determined.

33 Broad (1925, 118) established this standard philosophical usage.

34 Ironically, Dennett is a strong advocate for this kind of point about epiphenomenality. He has argued (1991, 401 ff.) that the most central meaning of the term “epiphenomenal” (as it was introduced by Thomas Huxley (1874) and is used in the psychological literature, for example) and indeed the only one that really makes sense, is just a non-functional property, one that has no effect on a certain process of which it is a by-product. The whistle of a steam engine, or doodling while in deep thought are, for him, paradigmatic examples.
Yet even if all this is so, it is still not clear that we have achieved a reductio of the zombie intuition. One might still want to say that, if it turns out that’s the way it is, then that’s the way it is: we cannot just insist that qualia physicalism must at all costs be true since such epiphenomenality could never be. Further, it so happens, we have at our disposal (as a result of earlier work in this thesis) a new way of deflating the ‘alienness’ objection—or at least casting it back upon an already well-known kind of ‘alienness’ in phenomenal consciousness (qua a seamless sequence of experiences).

Suppose phenomenal consciousness is epiphenomenal in the sense that it has no causal powers at all; it is only an effect, and not a cause of anything at all. Then it is not at all surprising that consciousness should be the only thing in the universe identifiable (by us) as having that property—indeed it is to be expected. For qualia are unique (as far as we know) in being such that to possess the property just is to be aware of the property (in the sense worked out in previous chapters); it is (probably) one of the few kinds of property in the universe that does not need some causally stimulated detector—such as a sensory organ or scientific instrument—to mediate our becoming (phenomenally) aware of it. Thus, it is possible to become aware of one’s own phenomenal states—that is, in fact, part of what it is to possess those states—even if those experiences have no causal powers, but this is not true of any (other) physical property we possess.\(^{35}\) Hence it is predictable that it is the only kind of property currently on the books which is a candidate for this radical kind of epiphenomenality.\(^{36}\)

\(^{35}\) This also seems to answer another standard objection to the epiphenomenalism of the mental, viz. “if qualia have no causal effects then how can we possibly know about them?” Note the vestigial oddness, though, in that that this would mean our beliefs about qualia—our propositional awareness of our own qualia (see Chapters Four to Six above)—are not brought about by the qualia themselves. Indeed, if we were zombies we would have exactly the same beliefs (or at least, beliefs\(^Z\)) about ‘our qualia’ even though we would have no qualia, and these zombie beliefs\(^Z\) would presumably be caused in just the same way as our actual beliefs.

\(^{36}\) The possibility remains open that the universe contains other kinds of epiphenomenal phenomenal properties that we do not instantiate in the right way and so have no reasonable prospect of coming to know about through science or philosophy—it is even totally impossible to imagine what it would be like to instantiate them. The only way such properties could be detected—since, \textit{ex hypothesi}, no measuring device could ever pick them up—would be through their phenomenal instantiation in us or, possibly, through extrapolation to it as the causal effect of
The second generator of discomfort with epiphenomenalism is based on evolutionary considerations. It apparently follows from the epiphenomenalism of qualia that qualia have absolutely no positive effect on fitness; they do nothing to render the animal more able to survive and prosper since, *ex hypothesi*, a creature exactly like it but lacking in qualia is in precisely the same position *vis-à-vis* the environment. There is therefore no evolutionary reason for the trait of having qualia, the argument runs, and, since evolutionary reasons are the only plausible ones in this area, no reason for qualia at all.

Indeed. But this is only a *reductio* of epiphenomenalism if there must always be functional reasons for a trait; and this is not so. For example, evolutionary reasons can often be *indirect*: trait $X$ (which has no adaptive advantage, and might even be disadvantageous) is constantly conjoined with trait $Y$ (which is sufficiently strongly adaptive), and so selection for trait $Y$ has led to the emergence of trait $X$ in that population also. Well known examples include the heavy weight of a polar bear’s coat (associated with its thermal value) or the pinkness of flamingos (consequent upon the type of food in their niche). Perhaps a similar story could be told about epiphenomenal qualia.\(^{37}\)

Further, the theory of natural selection is also perfectly compatible with the emergence and persistence of more-or-less neutral traits: traits that do not substantially harm an animal’s fitness, even in the sense that they occupy a ‘space’ that might be better filled by an advantageous trait. Such traits might be produced by random genetic variation, like any other, and then neither selected for nor weeded out by environmental pressures.\(^{38}\) Maybe epiphenomenal qualia could

---

\(^{37}\) In fact, given the thesis of even M1 supervenience, arguably it must be told. Clearly certain tendencies towards behaviourally relevant brain states are of adaptive value and these states *de facto* determine certain supervenient qualia.

\(^{38}\) This does however suggest that the capacity to experience qualia would have to be genetically determined and transmitted—it is interesting to think about what both the acceptance and the denial of this thesis would imply for the non-physicalist.
be construed as such a trait.

All of this goes to suggest that the appeal to epiphenomenalism does not in itself reduce the zombie intuition to absurdity—does not in itself show that zombies could not possibly exist. On the other hand, there is currently little substantial reason to take the predictions of the zombie intuition as true. Whether or not zombies are possible, and physicalism true or false, depends on the nature of qualia’s dependence upon the brain.

3. Conclusion

The metaphysics of qualia, then, revolve around the issue of dependence—and in particular, as we might put it, whether qualia are somehow ‘over and above’ the physical. Examination (under the rubric of supervenience) of the determination of qualia by the brain has brought us some useful results, and shown us something of what it would be for qualia to be consistent with the thesis of physicalism. The decision whether qualia physicalism actually is true or not, however, must wait for more empirical information about the dependence of qualia. The other major metaphysical question raised at the start of Chapter Seven—how does the ‘phenomenality’ of qualia work—must also presumably wait for more information on qualia dependence.

What further progress can we make now? Are there other considerations we can raise which might restrict the empirical possibilities for the dependence relation between qualia and states of the CNS? A standard a priori assumption about the metaphysical relation between qualia and the brain is that qualia are “at a higher level” than the brain’s basic physical properties. However, analysis shows that there is more than one way of being a “higher level” property—and that this makes a metaphysical difference, especially with regard to the prospects for reductionism and as to whether qualia are really, in any interesting sense, ‘over and above’ the physical. It will be useful, therefore, to get clear about the notion of “levels of description,” which I shall try to do in the following chapter, and then apply this discussion to the prospects for the dependence relation between qualia and the brain.
Chapter 9: Levels of Description and Explanation

Qualia are sometimes said to be “emergent from” the brain or “at a higher (or different) level than” that of the brain. These kinds of claims are often thought to accommodate the intuition that qualia, though perhaps ‘real,’ are just not ‘there’ at the level of sodium ions, axons and ganglions—they exist ‘only’ at some higher level. That is, the claim that qualia are ‘higher level’ properties is sometimes intended to express the belief that qualia are ontologically irreducible to neurones, but are nevertheless perfectly real and potentially physical properties. In this chapter I will examine what might be meant by these levels claims and why this kind of claim might be made. In the next I shall examine their implications for qualia. For brevity I will refer to such claims as invocations of hierarchies of levels or more briefly and more usually invocations of levels (and occasionally, for stylistic variety, I will merely refer to “levels talk”). These phrases should be viewed as terms of art or at least as semi-technical terms. They are deliberately non-specific: as will emerge, what I am talking about here is varied in kind.

I will argue in this chapter that invocations of levels are at once more complex and simpler than is standardly recognised: more complex in that there are a variety things that someone might be claiming and a variety of reasons why they might be claiming it; more simple because once we sort out the What and the Why, both will usually turn out to be fairly straightforward.

1. THE NEED FOR A GEOGRAPHY OF LEVELS

When philosophers, in my terminology, “invoke levels,” they usually, I think, expect themselves to be taken to be invoking some vertical scale of possible descriptions or explanations or beings. Very often, though not always, this spectrum is implicitly taken to be relative to an absolute “bottom level” of “full causal description,” with higher levels of description giving in-

---

creasingly less close approximations to that full description, causal explanation or basic being as you go up the scale.

These ways of thinking about what is involved in the invocation of levels are simplistic. The analytic technique of invoking levels tends to be treated uncritically in the literature, which seems to suggest that the technique is viewed as being monolithic.\(^2\) I will show instead that hierarchies\(^3\) of levels can be invoked for a variety of reasons and that the function of such invo-

---

\(^2\) The fact (or what I will argue is a fact) that we can invoke levels in a variety of ways is not widely recognised. In the field of the philosophy of mind, if people working in the area acknowledge the multiplicity of kinds of levels talk, then they do so primarily and merely by way of a contrast, which they take to be reasonably clear, between epistemological or perhaps pragmatic reasons to ascend to higher levels, and reasons of ontology. They tend to be primarily concerned to exclude (or, alternatively, defend) such an ascent where the ascent is for reasons of ontology: such moves are a focus of controversy. Meanwhile, most feel perfectly comfortable with such ascent when it is “merely” for epistemological or pragmatic reasons, tending to talk in terms of the in principle descent to lower levels. My analysis here suggests that these assumptions need to be challenged more critically.

\(^3\) I will loosely call them “hierarchies” for now since the familiar term makes plain what I mean at this point. However, we shall see, not every species of levels-structure is hierarchical: in particular, explanatory levels do not form any clear hierarchy. (I shall sometimes use more accurate term “structure” below.) Why, then, speak of “levels” at all? Why not just cash the concept out in terms of regularities visible “from a certain point of view” or “from within certain explanatory frameworks”? Doing so would have the benefit of divorcing one from the dubious presuppositions that:

a) the visibility of equivalence patterns is correlated with levels of detail or generality, and

b) there is a hierarchy of levels, falling into an ordered pattern and going down to a “base level.”

Both these assumptions seem open to criticism. For example, on the first point, geometrical regularities are visible at all levels of detail; and within “levels”—such as molecular physics for example—there is a perfectly reasonable sense in which much more or less detail can be gone into. On the second point, the notion of a “fundamental level” seems really quite bizarre upon consideration; further, many “levels” do not fall easily into a hierarchy—should the level at which ethics is studied be considered higher or lower than the level of economics, for example?

Nevertheless, since this is really merely a matter of terminology, I would like to continue to use the word “levels” (in order to tie this discussion in with the relevant literature), but to divorce it of—or rather, render it neutral with respect to—the problematic associations outlined above. My own view of the matter is something like this: It seems to me that in fact the different descriptive lexicons do in fact fall into line generally with levels of generality. The problem of the “fundamental level” is one I am highly sympathetic to, but I’m not sure it need be a real problem—the request for a “bottom level of description” need make no more sense than that for a “maximum level of detail.” (However, here I side-step worrisome questions to do with “the level of causation.”) And I do
cation can similarly vary. Moreover, I will argue that the extent to which the invocation allows for reducibility or eliminability of the higher levels in favour of the lower also varies. Correspondingly, talk _simpliciter_ of the role of such invocation or the reducibility of the various levels in such an invocation can never be accurate. There are in fact, I shall suggest, at least five distinct breeds of hierarchy of levels of description: three sorts of epistemic/pragmatic use, and two potential ontological uses.  

What I am providing here is not a “true theory” or “correct description” of the invocation of levels, for one of the things I in effect argue is that there can be no such _single_ theory. Instead, I aim to provide a conceptual framework for the invocation of levels, for levels talk, a framework in which it will be clear that in levels talk one may be saying a variety of things and arguing a variety of claims. Different uses of different hierarchies of levels, once so recognised, can be seen to be differently motivated, so that we can, for example, simultaneously accept one kind of ascent up levels in a certain domain whilst finding another perplexing, in the same domain. Similarly, one might argue that a particular hierarchy could be of more than one kind; serving, that is, more than one purpose. Thus, for example, one might well find that a certain ascent might be simultaneously epistemologically convenient and explanatorily necessary, and that higher levels are reducible in one way but not another.

---

4 This is a taxonomy, or logical geography, not of levels of description themselves, but of _hierarchies_ (or structures) of such levels. For example, a particular level (such as the neuronal level or that of fairly high-level functional organisation) appears in quite a different light depending on whether it is, say, part of a hierarchy where the higher levels are merely descriptive shorthand for the lower, or instead part of one picking out degrees of causal emergence.

5 Thus, for example, when people say that qualia are emergent entities or entities existing at higher levels than the brain they are not always making the same claim, let alone making it for the same reasons. Some may be making a metaphysical claim, a claim about what qualia _are_, or they may be making a claim about what we need to do if we are to understand or to explain the nature of qualia. Not all such claims, then, are equally substantive or defensible.
2. EPistemological AND PRAGMATIC LEVELS

The primary distinction to be made is that between epistemologically or pragmatically motivated levels on the one hand and those demanded by ontology on the other. Within the epistemological camp, it turns out that there are at least three sorts of levels hierarchy sufficiently distinct to count as three different types: hierarchies constructed for reasons of what I call epistemological convenience, modelling, and explanatory pragmatics. Within the roughly ontological group there are two distinct varieties: group-level properties and, if they exist at all, what are sometimes called “levels of being.” I shall deal with the epistemological/pragmatic species of levels talk first.

a) Levels of Convenience

This is the most evident type of invocation of levels: higher levels that are used because it is in some way too cumbersome to give full descriptions or explanations or accounts at lower levels. Examples of this kind of invocation of levels are commonplace. It is (I take it) perfectly possible in principle to truly describe the operation of a car over the span of, say, ten minutes, in terms of the interactions of all its different parts within the framework of general relativity theory; however, it is far easier and almost always sufficiently informative to tell the same story in the language of spark plugs, carburettors and cam shafts within a sort of ‘folk Newtonian’ causal theory. A complete historical account of the battle of Agincourt, again, would speak in terms of the movements of groups of men, of general strategic and political considerations, of the beliefs and attitudes of Henry V and Constable d’Albret and so on; it would not attempt to give a description or explanation or account at the level of individual archers, let alone in terms of the physical equations that governed the flight of each arrow.

When we ascend to higher levels for reasons of convenience, the ascent might be for reasons, as we may say, of mere convenience: it is just quicker, less irritating or less tiring to work at the higher level, though we could descend to lower levels if we were not so lazy. In other cases, however the convenience is not mere convenience but because work at the lower levels is actually a practical impossibility: there are situations where we just could not, at the lower levels, hold
everything in memory, or compute things fast enough to get any results in a human life-span.⁶ To complicate matters further, if we find it impossible to work at the lower levels, this could either be a constraint contingent upon our current technological development (the relatively paltry power of current particle accelerators, for example, which is apparently inadequate by an order of magnitude for testing the claims of string theory), or might be, for some reason, in principle impossible for us to overcome (just as, perhaps, it is in principle impossible from any given standpoint to measure the simultaneous position of every particle in the universe, since most observable parts of the universe are, so to speak, not at time \( t \) but at time \( t - n \)—are behind us in time). Thus ascent to higher levels for the reason of convenience is sometimes just not dismissible as “in principle eliminable,” as some have labelled it, at least in the sense that it need not be “in principle eliminable” by any human being—or, in some cases, any finite describer at all—no matter how conscientious and with how much time on their hands.

However, this kind of use of levels of description is fully consistent with (though it does not entail) the view that there is only one fundamental and complete description or explanation or account of the physical universe (at the level of ‘fundamental’ particles and forces), and that descriptions or explanations or accounts at the higher levels are ontologically fully reducible to, or even eliminable in favour of, those at the ‘bottom level.’ Nothing I have said above contradicts that usually physicalist assumption; the claim that the city of Calgary’s infrastructure is vastly complicated, for example, does nothing on its own to suggest that it is not at bottom made up of quarks and fields.⁷ In summary, levels of convenience are consistent with the ontological reducibility or eliminability of all higher levels to the most basic, but may involve a kind of pragmatic irreducibility that can be almost as serious as the irreducibilities of the other kinds

---

⁶ The highly complex systems described by dynamic systems theory, such as this planet’s atmosphere, are popular examples in this area; but any sufficiently complicated chunk of reality will do. Though the behaviour of, for example, all the fixed objects (roads, bridges, wiring, buildings, and so on) in the city of Calgary is fairly predictable, a bottom level description of that collection at time \( t \) would presumably be at least as difficult and complicated as that of, say, a small storm.

⁷ Indeed, the assumption that macro-level objects are divisible into vast quantities of smaller and smaller component parts is probably a large part of the intuition supporting their complexity.
we will examine.

b) Model Levels

To use a model to explain or describe some phenomenon, the model explaining or describing the phenomenon in a particularly helpful way not available for some reason at the lower level, is at least in some cases to move from one level to a higher level. For example the more abstract structural level of a model might have greater heuristic power in formulating testable hypotheses or theorising about the connections between parts of a domain: AI-like computer programs can be (and often are) treated as providing models of human cognitive processes, whereby the program’s algorithms and the symbolic entities they manipulate are seen as high-level analogues of elements of actual human cognition. Examples in other fields might be the planetary model of the atom or the person-as-agent model of human action.

The reducibility of models to lower levels is a complex issue on which I have little new to say and so shall not dwell. It is consistent with model-style invocation of levels that, at a particular time, they need not be ontologically reducible to lower levels, and need not be eliminated if they are not so reducible. Further, that two models within the same “theoretical space” are inconsistent with each other is not itself reason to reject one of them. The nature of the irreducibility or ineliminability in these cases is a function of the fact that the goal of the modelling is not directly the representation of ‘reality’ but the search for an instrumentally fruitful structure for information about the domain.

Models are sometimes taken to be actually true descriptions or explanations or accounts of the domains they describe or explain or account for: for example, a chemist might (once have) come to believe that the planetary model of the atom is in fact a literally accurate description.

---

8 Smolensky 1980 is a central example of this kind of invocation.

9 Identifying this kind of levels talk relies on identifying the activity of providing a model for some domain. It may be that models, and the associated levels of description, are related to each other not by satisfying necessary and sufficient conditions, but by being “members of the same family.” For my points here, I do not need to have a view on this.
of atomic structure. When models are treated in this ontologically robust sense then there are three possibilities:

a) The chemist may be wrong, and the model is in fact of the instrumental sort described above. This is the core type of a model level of description, explanation or account.
b) The chemist may be right, and furthermore the properties at the level of the model are ontologically reducible to properties at the lower levels: in this case, either the description, explanation or account ceases to be a “model” and becomes a literal description, explanation or account, or the model level continues to be used simply because of its heuristic/pragmatic usefulness, as a level of convenience and/or as an explanatorily necessary level (see the next section).
c) The chemist may be right, and furthermore the properties at the level of the model are not ontologically replaceable by properties at the lower levels: in this case, the model level has to be retained because it contains ontologically emergent properties. This kind of level is discussed below.

c) Explanatory Pragmatics

This, in my view, is the most interesting and under-recognised of the varieties of epistemological invocations of levels. This kind of invocation of hierarchies of levels arises from the nature of explanation itself; in order to construct good explanations we are often required to operate at higher levels of description or explanation or account. There are certain patterns in the world that must be picked out in performing certain explanatory tasks and which are only “visible” at certain levels of description or explanation or account; this has often been commented on before, but what is not always noted is, I contend, that such patterns are only available (or appropriate) for use in these explanations when described by higher-level predicates. Thus, when one’s aim is to make good explanations, these higher levels are in principle not eliminable.

---

10 Such as in Dennett 1987 and Pylyshyn 1984.
Consider Zenon Pylyshyn’s well-known “9-1-1” example.\(^{11}\) Suppose you are standing on a street corner and observe a sequence of events that can be described as follows. A pedestrian, who is walking along the sidewalk, suddenly begins to cross the street, causing a car to swerve and crash into a telegraph pole at the side of the road. The pedestrian runs to a nearby telephone booth and dials the numbers 9 and 1. What explanation could we give of these events? How could we predict and explain what the pedestrian will do next? Pylyshyn uses this example to try to show that intentional terms are somehow irreducible\(^{12}\); but that is not my concern here. I want to put a different spin on the example and suggest that it could be used to illustrate the claim that such explanations can only be adequately given in the descriptive vocabulary of something like folk psychology because any explanation using lower-level predicates would simply be a poor explanation: it would lack many of the epistemic and pragmatic virtues that are required of good explanation. And this is true not just of psychological explanation but also a wide variety of other higher-level explanatory projects, such as economics and epidemiology. Important to note here is that the claim is not merely that such explanations are just easier at higher levels of descriptions, but that anything at a substantially different level of description could not possibly count as the explanation required.

The reason for this is that the explanation relation\(^{13}\) is not wholly a semantic relation, but is a substantially pragmatic one.\(^{14}\) It is not (unlike, perhaps, the truth relation or the relation of empirical adequacy\(^{15}\)) a relation that holds simply between, say, a theory or hypothesis and the

\(^{11}\) Pylyshyn 1984, 3 ff.

\(^{12}\) That is, to reinforce a point made by Davidson (1970), Dennett (1978), Fodor (1980) and Putnam (1960, 1973), among others.

\(^{13}\) I.e. the relation of x’s being explanatory of y.

\(^{14}\) I take “semantic” to mean something like “having to do with the meanings of sentences, and the logical relations between them,” whereas “pragmatic” means something like “having to do with factors extrinsic to the meaning and logical structures involved, such as context and aims.” The former deals with ‘language’ in itself; the latter with the actual use of an actual piece of language.

\(^{15}\) Van Fraassen has gone so far as to argue that “there are no explanations in science” (1977, 225). Indeed, he suggests, a theory need not be true or even accepted to have explanatory value; rather, the explanation relation is visible before acceptance, and is part of what leads us to accept the theory. Further, a theory known to be false can
Chapter 9: “Levels of Description.”

phenomena or facts to be explained; rather it depends additionally upon various contextual features. Since it would require at least another chapter to argue properly here for the (not especially controversial) claim that explanation has or can have a large pragmatic component, instead I will briefly put forward four reasons or pieces of evidence in its support.

First, the function of explanation is to increase understanding, and pragmatic, contextual factors (not semantic ones) operate to determine which are the “epistemic gaps” that need “filling.” As Michael Scriven puts it, “The meaning of [the concept of explanation] springs entirely from the ‘psychologistic’ notion of understanding for which it serves, in various circumstances, as a carrier” (1975, 9). For example, in answer to the question, “Why is the porch still be explanatory in certain areas; Newtonian physics, for example, explains the tides. (See Van Fraassen 1977, 217–18.) Nancy Cartwright (1980) also argues that the roles of ascertaining what is true in nature and of how we are to explain it are entirely different functions and should be kept distinct. To describe the world, she urges, is not to explain it. Nearly every quantitative law in physics is a ceteris paribus law: they hold only in special circumstances, often quite infrequent ones. One of Cartwright’s examples is Snell’s Law (describing the angle of refraction and incidence for a ray of light) which is usually given as a universal generalisation, and yet which holds only for an interface between two optically isotropic (i.e., exhibiting the same optical properties in every direction) media; that is, only in ideal cases. Further, there are usually in science no laws to govern the combined effects of more than one law: there is no principled way to combine, for example, Fourier’s Law for heat flow and Ohm’s Law for electric current (when there is both a temperature gradient and a potential difference at work). Thus, Cartwright claims, the laws of nature cannot be our grounds for saying that a covering law account is an explanation, for we do not know these laws: we know only rough approximations to them. Yet ceteris paribus laws do explain: generalisations do explain results compatible with their predictions, even though they are not literally true.

16 See, for example, Scriven 1975 and Van Fraassen 1977. “Which factors are explanatory is decided not by features of the scientific theory but by concerns brought from outside. This is true even if we ask specifically for an ‘efficient cause,’ for how far back in the chain should we look, and which factors are merely auxiliary contributors?” (Van Fraassen 1977, 224). Arguably, Aristotle recognised this in his division of science into demonstration (in the Posterior Analytic) and explanation (dealt with mainly in the Book II of the Physics). His doctrine of four causes can in this way be seen as allowing for the systematic ambiguity of why-questions.

17 Here is a summary of how Scriven 1975 cashes out the notion of explanation:

a) Explanations are symbolic vehicles for conveying understanding.

b) Understanding is acquired whenever the capacity for solving a certain appropriate range of problems is achieved without learning the solutions for each problem separately.

c) Understanding may thus be acquired by all sorts of inputs, scientific or not, depending simply on whether or not they achieve this result.
light on?” the answer may be (among other things) either “Because I flipped the switch” or “Because we are expecting company,” depending on the context. Possible answers to the question “How do I get to the party?” include “Come with us,” “Go down the hill until you hear the noise,” or “It’s at Ann’s house.” Which is appropriate and which not depends on the context. Another example which is less apparently psychologistic, is the explanatory relationship between a pole and its shadow: for a flagpole, clearly the height of the pole explains the length of the shadow; but on a giant sundial, the properties of the shadow may count as explaining the pole’s length (e.g. why we needed such a long pole).

Second, good explanation must make salient the relevant explanatory information to the intended audience, and what is relevant largely depends upon context—for example, salience is highly contingent upon the beliefs the intended audience already happens to hold. A narrative that parenthetically includes a large amount of irrelevant information is a bad (confusing) explanation; one that fails to distinguish between the relevant and irrelevant information ceases to be an explanation at all. In addition, an explanation will frequently be required to deal with a range of possible cases, and not just the one under consideration. That is, to count as an answer to a particular demand for explanation, an explanation must often not simply describe the specific causal chain under consideration, but must abstract out and make salient the factors which explain the general phenomenon.18

We explain the fact that a square peg won’t pass through a circular hole, for example, by the general fact that squareness won’t pass through an equivalent area of roundness; we do not provide the ‘real’ causal story at the subatomic level—we don’t give the full story of all the actual, causally efficacious factors. The reason for this is not primarily that we have epistemological problems dealing with all those quarks,19 but that the subatomic facts are not the ones we wish

---

18 Andy Clark (1989, following Jackson and Pettit) calls these “program explanations,” as opposed to “process explanations” which, by contrast, narrate the very features which are causally efficacious in a particular explanandum.

19 Though this has also been put forward as a problem for explanation. William Seager, for example, has argued that explanations must be “intelligible,” where by this he means that it is not enough that “if one had enough
to group together. At the subatomic level, roundness and squareness are not “visible” but it is square pegs and round holes for which we wish to formulate a general principle. Further, there might be some Twin Earth where matter is not mesons and bosons but Xs and Ys; there, the causally efficacious micro-structural facts will be different, but still higher-level descriptions and explanations in terms of roundness and squareness will be operative. So, the argument runs, such explanations, if they are intended to have the correct sort of generality, cannot be phrased in terms of mesons and bosons. This is not because there is no other, more fundamental, causal story to be given; we need not deny causal closure to insist on higher-level explanation; the problem is that a causal explanation would capture the wrong sort of generalisations.

Third, the explanation relation has a particular kind of asymmetry which renders it non-equivalent with such logical or semantic relations as constant conjunction and deduction. Falling barometers are constantly conjoined with approaching storms, but the explanatory relationship runs only one way: the occurrence of either one is very good evidence for the other, but this (symmetrical) evidential status is not enough to qualify falling barometers as being explanatory of storms—there must be something different about the two “directions of explanation,” and this difference cannot be one of deducibility. Similarly, the fact that the higher-level behaviour of a system can be deduced from its description as a system of elementary particles would not imply that this behaviour can be explained by that description. Think again, for instance, of Putnam’s example of square pegs failing to fit into round holes (Putnam 1973): talk-
ing about clouds of elementary particles would fail to properly explain the phenomenon. That the behaviour could be deduced from the properties of the particle fields is not the point; as Putnam puts it, this is like deducing $F$ from a complex mathematical transformation of $G, H$, such that the information $G$ is practically impossible to recover from $H$, when $F$ is deducible simply from $G$ alone. This is not an explanation.

Lastly, explanation is not a transitive or acyclical relation, and therefore much of the information supplied by low-level descriptions is not relevant to the explanation. Barometers will fall if a storm is coming, and the coming of the storm depends upon atmospheric conditions several (or even several hundred) miles away, but typically atmospheric conditions several hundred miles away do not count as explaining the behaviour of barometers—this is an example of the non-transitivity of explanation. Likewise, explanations of some higher-level phenomenon $B$ in terms of a lower-level one $A$ are not part of explanations of some still higher-level phenomenon $C$. A complete explanation for the fact that a peg will not pass through a round hole is supplied by the fact that the hole is rigid and smaller than a section through the peg. Further information may also tell one why the peg and board are rigid but this is extra; it is not part of the explanation (though it may certainly help to explain that explanation). To explain the rigidity is not to explain why the peg will not fit; it is the rigidity itself which explains that and the micro-information in itself only explains that which explains the explanation.

The important point to emphasise for my purposes is that, due to its pragmatic nature, there are certain virtues that an explanation must have, such as simplicity, generalisability, relevance, brevity, salience, ease of use and understandability, and these virtues are frequently unavailable to those working at the lower levels of description. In certain contexts, an increase in the generality of explanations, though bought at the cost of becoming a less full account of the particular causal chain responsible, is nevertheless an explanatory virtue—indeed, a necessary condition for explanation to take place at all. Thus, in order to explain effectively we must nec-

---

21 That is, if $A$ explains $B$, and $B$ explains $C$, it does not follow that $A$ explains $C$.

essarily often use higher levels of description: we cannot, for example, replace folk psychological explanations with adequate neurophysiological explanations. To respond to a request for an explanation of why the pedestrian dialled 9 and then 1 with a story about quarks, gluons and weak atomic forces would be inappropriate—it would be at best an extremely poor explanation, if we were to count it as an explanation at all (quite likely, it would utterly fail to fill the relevant “epistemic gaps” in the listener’s cognitive structures). Similarly, good explanations of everyday physics, the weather, and why aeroplanes stay in the air could not be made at the bottom level. And all this is true regardless of the reducibility of the higher level descriptions to the lower level ones.

Though superficially similar to the levels of convenience species of levels invocation, this explanatory pragmatics kind is importantly different. Presumably, a sufficiently powerful god-like being would have no need to construct a hierarchy of levels of epistemological convenience; but even such a god would be in principle unable to provide for us, at the ‘bottom level,’ a good explanation of many perfectly real phenomena. It is not even clear that a god could construct certain good explanations for another god at the bottom level, again because it is one of the virtues of a good explanation that it is no more complex than necessary, and that it makes salient exactly what is relevant.23

Further, explanatory pragmatics give rise to a different kind of structure of levels of description. Levels of epistemological convenience are probably hierarchically nested: they presumably form a series of levels from most complex to least complex, and the degree of complexity of any one level can reasonably be compared, at least roughly, to that of any other level of description. By contrast, explanatorily pragmatic levels often do not appear to be ‘above’ or ‘below’ each other: the level of economic explanation and that of biological explanation at the level of species differ, but it is a fundamentally moot point which is the ‘higher’ level.

23 This difference makes clear one important part of the distinction between description and explanation: it may turn out to be in principle possible to give an ontologically complete description of the universe at only one ‘bottom’ level, but it will never be that all explanations could use that level of description—they will have to avail themselves of higher levels of description.
Finally, although, as with epistemological convenience, we are dealing with a pragmatic notion here, the stakes have been raised a significant notch. Suppose we wanted to construct a science dealing with the everyday behaviour of human beings, but that, by some mysterious and implausible chance, we had developed no vocabulary at the level of folk psychology (only, let us suppose, Newtonian physics). In such a case, we would find it *in principle* impossible to give good explanations of everyday human behaviour without moving to—inventing new predicates at—the folk psychological level.

3. ONTOLOGICAL LEVELS

So much for the classification of epistemological/pragmatic hierarchies of levels; now on to ontological levels claims.

*a) Group-Level Properties*

Though it is an open and (deservedly) controversial issue whether there are what might be called “levels of being,” there is one sort of roughly ontologically motivated hierarchy of levels of description which seems fairly uncontroversial, albeit toothless. It is a logical commonplace that groups may have properties which are not properties of any or all of their components: for example, a soccer team has the quality of being made up of exactly eleven human beings, but none of its components have that property. There are a wide variety of such properties: having to do with the constitution of the group in question (e.g. having eleven parts, being octahedral), its relations with other things (e.g. being larger than a bread-box), its function (e.g. being able to run a country, being a can opener), and so on.

For this kind of property or pattern, a higher level of description must be adopted. Isolated examination of each of the constituent parts of the group in question, no matter how detailed, will not reveal these group-level properties; it is necessary to consider the group as a mereological whole. So, insofar as it is true that soccer teams have eleven members, and insofar as it is true that none of the members of a soccer team have that property, and insofar as it is
true that soccer teams exist, then a full description of the universe requires us to ‘stand back’ and consider certain blocks of basic materials as if they were individuals: as property-holders in their own right.  

However, this kind of ontological level-making, in itself, is rarely seen as a serious motivation for belief in “levels of being” (as we might put it using a term of art used by C.B. Martin (in conversation))—the belief, roughly, that the universe compartmentalises itself into different levels such that upper levels are ontologically or metaphysically held apart from lower levels, probably through being irreducible to those lower levels. The reason why mereological levels, as we might call them, are unthreatening is because they are consistent with one or both of the following moves: one can either deny that particular group-level properties are really “ontologically serious” (whatever exactly we might take this to mean) or one can show (or at least hope to show at some point in the future) that group-level properties are fully describable at the ‘bottom’ level, the level of reality. An example of the former kind of move would be to say that being a soccer team is not a scientific property: the fact that such a property would not feature in a full description of the universe is not a problem, because the full description is supposed to be a scientific one, including only “real,” “ontologically serious” properties. An example of the latter kind of move would be to proffer a promissory note to the effect that the predicate being a soccer team is fully describable at the base level in terms of, perhaps, the physical make-up of the players in combination with the bottom-level physical descriptions of all the relevant societal beliefs, attitudes and institutional structures which go to make something a soccer team (rather than, say, attendees at a board meeting).

As a tactic I prefer the second move; the first just seems to raise too many difficulties in dividing up the properties in the world into “real” and “non-real” in a non–ad hoc way. However, needless to say, it would be hugely difficult to actually carry out in practice the project of reducing mereological predicates, even just a few examples to show that it can be done: it is

---

24 In fact, it is possible and perhaps even obligatory to treat virtually all predicates, except a few like strangeness, charm and being an up-quark, as being group-level.

25 It is not clear what exactly this comes to, a point I shall enlarge upon somewhat below.
probably a promissory note that is destined to remain forever unpaid. On the other hand, adopting such a reductionist picture shows how deflated are the ontological claims of mereological hierarchies: it makes some intuitive sense to continue thinking of group-level properties as being at a “higher level” than the properties of their components, but in fact, on this picture, they are really just very complex concatenations of so-called “bottom-level” properties.

b) Levels of Being

So what would it be to assert the existence of levels of being? A starting point might be that there are two conditions:

1) Descriptions at some higher level must be irreducible: that is, they must capture patterns or generalities which cannot, even in principle, be captured by descriptions at lower levels.

2) The patterns or regularities so picked out must be ineliminable: they must be necessary for a full scientific description of the universe … must be ontologically real, rather than merely instrumental or apparent.

What would it be for a pattern to have these attributes? I will have most to say about the first condition here. One common approach is to talk about patterns describable only at something like “the right level of detail,” where what is salient about a range of cases can be picked out. For example, Pylyshyn argues that the level of functional architecture is “the right level of specificity (or level of aggregation) at which to view mental processes.”26 This is because the functional architecture level is considered by Pylyshyn to be the only level on which what he considers the relevant regularities in computation—the algorithms operated by the brain, no matter how they are physically realised—can be spotted. What Pylyshyn seems to be alluding to here as the reason why particular patterns are visible only at certain levels is the notion of multiple realizability (although he does not put it in those terms). Similarly, for Hilary Putnam, the multiple physical realizability of mental states casts doubt on the prospect of any bridge law

26 Pylyshyn 1984, 92.
linking physical and mental predicates. Many non-physical predicates (such as being a spark plug or ordering 500 shares in General Motors) may be constituted by an indefinite number of physically quite different things. Hence, for many upper level predicates, there can be no bridging generalisations correlating them with a specific kind of physical event. Concepts at one level cross-classify those at the other level, so that there is no way to identify properties at one level with those at the other.\(^{27}\)

However, as is now widely recognised, multiple realizability by itself is not enough to get any strong form of irreducibility: though higher levels may be multiply realized, it remains possible that they can still be explained by, and described in terms of, whatever structure does realise them. Further, complex predicates, including statements of initial and boundary conditions, might in principle be constructible at the bottom level which correspond to upper level predicates. Put another way, though there is no type-type reductionism, token-token reductionism remains, and in principle a lower-level predicate can be constructed from the disjunction of labels for all the lower-level tokens.\(^{28}\)

I think it is much less widely recognised, on the other hand, that multiple realizability can be slightly added to in such a way that it does bring about strong irreducibility of higher-level patterns. This happens if we simply add the further stipulation that there is no relevant similarity or regularity existing between, and only between, the group of token physical realizations of an upper level pattern; that is, that there is no lower level pattern corresponding to the upper level pattern. The pattern is realized, but the set of tokens which realize it have nothing uniquely in common. Then the upper level regularity, qua regularity, is not reducible to the lower level of description: the regularity does not exist at that level. The upper-level property is not that property (or indeed any property) at lower levels.

I am not aware that science has yet definitively discovered any examples of such an irreducibly high-level pattern (though possible candidates might be being objectively scarlet or being a

\(^{27}\) See Hooker 1981, 496, on cross-classification. Dennett 1995c claims that processes, and in particular algorithms, are also level-dependent, while Haugeland 1982 applies the notion to events.

\(^{28}\) See Fodor 1974.
certain Mendelian trait), but perhaps a simplified example of something that is not like this might make the contrast clearer. The property of liquidity has, let us suppose, a set of satisfaction conditions at the level of ordinary observation: something is “liquid” if and only if it satisfies those conditions, such as being pourable or able to flow around solid objects. At a lower level of description, there is also a set of satisfaction conditions for being liquid: roughly, having molecules that move around each other in a certain way. Thus liquidity is a reducible property. Some property is not reducible if a similar story cannot be told.

It seems to me that this is a moderate, convenient and clear notion of what it would be for a pattern to be realized by lower-level elements but not reducible to them. Indeed, the claim that such patterns might actually exist has some prima facie plausibility. They would still be patterns of lower-level entities, not floating somehow metaphysically free, and perhaps need not violate the constraint of causal closure. Further, one can describe actual patterns which do seem to be good candidates for this kind of irreducibility: those relied upon by the special sciences, such as economics, are often cited cases, especially since they frequently describe regularities which are not causal regularities but are the combined distinct results of several different base-level regularities. Inflation, for example, it is often urged, is not a cohesive physical event which might instantiate some physical law: the pattern representing inflation, it seems prima facie reasonable to say, does not exist at the level of quantum mechanics. Analogously, the path of a single billiard ball is physically explicable, and subsumable under a particular lower-level regularity; the behaviour of all the billiard balls on the planet in the last five minutes is not, even if it formed some significant pattern (such as replicating the movements of the molecules of a certain volume of gas under particular conditions), because it consists in the mere agglomeration of independent causal processes.

On the other hand, there are also strong reasons to think that no serious ontological patterns will turn out to be irreducible in this way. In particular, I think we should be clear that the assertion of this sort of ontological irreducibility for a pattern \( \Phi \) would commit us to the claim

---

29 Owens 1989 makes this case.
that the token instances of that pattern do not have anything uniquely in common: thus either some tokens which are utterly different from some other Φ-tokens must themselves also be Φ, or others which are qualitatively identical with some Φ-tokens must fail to possess it, or both.\footnote{Consider this simplified example: Suppose some universe contains exactly eight objects (1–8), exactly three base-level properties (a, b, c) and exactly two higher level properties (Ψ, Φ). In the following example, Ψ is reducible but Φ is radically irreducible.} Irreducible patterns could be such that two utterly different base-level structures possess them, but also could be such that, of two qualitatively identical base-level structures, only one possesses it.

The latter situation would usually be taken as violating supervenience and becomes fairly implausible, I think, if one seriously considers all the properties of the instantiating tokens of a pattern, including all their manifest and non-manifest (but still actual) inter-relatednesses and inter-reactivities—it then becomes much harder to imagine plausible situations where these properties do not make the pattern what it is. That is, it becomes more difficult to hold coherently before the mind’s scrutiny the proposition that these tokens with these properties could have instantiated some other pattern, or no pattern at all.

Consider again the inflation example. Try (as best one can) to imagine the vastly complex combination of basic properties which goes to make up this pattern, including the physical instantiations of beliefs and attitudes towards money, the changes over time in the purchasing power of fiscal tokens (described in physico-causal terms), the (physical tokenings of the) beliefs people have that this is inflation, and so on. If it were possible to give such a ‘bottom-level’ description at all—and we have so far agreed there is no reason to think it is not possible—then it is difficult to imagine that it could be of anything other than inflation.

Similar considerations also weigh, though less forcefully, against the other possibility above; that some of the different token instances of the same pattern could have nothing relevantly in common. Any instantiation of inflation would surely bear some significant resemblance to any
other, with respect to, say, basic physical instantiations of the sub-patterns of having certain societal attitudes towards money and there being certain changes in purchasing power, just as any physical instantiation of a belief that one’s grandmother is in the room must have a subset of certain relevant dispositional powers to bring about behaviour in certain organisms and settings, certain connections to other beliefs about grandmothers, and so on.

4. Conclusion

At least four significant general conclusions seem to me to follow from these considerations. First, it seems to me an important and under-recognised fact that there is really no unified notion of invocations of levels of description, and that such levels talk can operate in different ways. Second, there is an interesting and little emphasised non-ontological motivation for in-eliminable levels which follows from the very nature of explanation itself and its pragmatic virtues. It is not only because of our human and technological limitations that we cannot spend our whole lives at the ‘bottom level;’ it is also in some cases because of the nature of explanation that this is so. Third, I have attempted to provide a clear, plausible account of what might be meant by the elusive notion “levels of being,” concluding that it consists in an absence of correspondence between the reference of ontologically serious predicates at the higher-level and any identifiable pattern or regularity at the lower. Finally, I have suggested that the assertion that there are levels of being corresponding even to this modest notion is less plausible than it might appear at first sight.

<table>
<thead>
<tr>
<th>Kind of Levels</th>
<th>Motivation</th>
<th>Reducibility</th>
<th>Eliminability</th>
<th>Structure</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epistemological Convenience</td>
<td>Convenience/processing limitations.</td>
<td>Consistent with ontological reducibility, but perhaps not epistemologically reducible.</td>
<td>Consistent with ontological eliminability, but perhaps not epistemologically eliminable.</td>
<td>Hierarchical (with indefinitely many levels).</td>
<td>Historical narratives, economic theory, folk psychology.</td>
</tr>
<tr>
<td><strong>Modelling</strong></td>
<td>Modelling a phenomenon to aid further research.</td>
<td>Consistent with ontological reducibility, but often in fact irreducible</td>
<td>Consistent with ontological eliminability.</td>
<td>Non-hierarchical (indefinitely many levels).</td>
<td>Planetary model of the atom, perfect market model of economy, person-as-agent model of action.</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Group-Level Properties</strong></td>
<td>Logical concerns about groups.</td>
<td>Consistent with ontological reducibility.</td>
<td>Ontologically ineliminable.</td>
<td>Hierarchical? (two levels?).</td>
<td>Having eleven members, being a swarm of bees, being able to govern a country.</td>
</tr>
<tr>
<td><strong>Ontological Emergence</strong></td>
<td>Needed for a complete account of the universe.</td>
<td>Ontologically (and explanatorily) irreducible.</td>
<td>Ontologically (and explanatorily) ineliminable.</td>
<td>Hierarchical (two levels?).</td>
<td>?</td>
</tr>
</tbody>
</table>
Having completed this analysis of the notion of “being at a higher level than,” the next (and, for us, almost final) order of business is to gauge, for each of the five types of level-hierarchy in turn, what would have to be true of qualia for them to be higher level properties in the relevant way. Our main interest is to assess the metaphysical or ontological consequences for qualia—especially with respect to the dependence relation between qualia and fairly low-level properties of the brain—and as we shall see such consequences follow fairly directly even for the pragmatic and epistemological types of level. At the risk of performing armchair neuroscience without a net I shall also make some speculations about which of these consequences are more plausible than others (with the tacit proviso throughout that plausibility is just that—at most, a reason for pursuing one research project rather than another—and is not a constraint upon the structure of the universe).

1. PRAGMATIC AND EPISTEMOLOGICAL LEVELS AND QUALIA

Though somewhat preliminary to our mainly metaphysical concerns, examination of the status of qualia within these three varieties of levels hierarchy does bring to light some relevant and interesting considerations. En route to claiming that qualia are probably at higher levels of convenience and explanation but do not feature in any modelling levels hierarchy, I shall define and discuss the important notion of ontological atomicity (which will be of use later), address the claim that qualia are merely theoretical constructs, and briefly revisit the issue of epiphenomenalism.

a) Convenience and Atomicity

What would it be for qualia (like virtually all of the referents of our everyday predicates) to be at a higher level of convenience than, say, basic physics or some future Theory of Everything (TOE)? Well, clearly, qualia will be (merely) at a higher level of convenience than some other
level only if they are not at the ‘bottom’ level of convenience. For qualia to be at the bottom level of epistemological convenience, let us say, they would have to feature in the (simplest possible) maximally complete, exhaustive and detailed description of the universe. This in turn would be the case only if a) the referents of qualia terms, unlike say demonic possession and phlogiston, actually do exist, and b) qualia terms could not be replaced, at the lowest level, with more complex but more exhaustive and detailed descriptions and explanations that do not mention such properties. I shall label this position the view that qualia are “ontologically atomic,” meaning roughly that they are not analysable into combinations of other (more fundamental) components. By this we do not mean just “unanalysable at a certain level of epistemological complexity (such as that of folk psychology),” but must mean unanalysable simpliciter, at any level even the most complex. The assertion has to be, for example, not just that I cannot provide any more detailed introspective analysis of my qualia than to say, for example,

---

1 I said above that levels of convenience are hierarchical and I shall assume that there is such a thing as a ‘bottom level’ of such a hierarchy. To be ‘at a higher level’ simpliciter, then, will make sense if interpreted as being at a higher level than the lowest—i.e., simply, not at the bottom level of convenience.

2 The “simplest possible” clause is supposed to rule out wilfully and unnecessarily complex (but still true) theories, if such are possible at the bottom level. The maximally complete true description (and I just assume for the sake of this argument that such a description is in principle possible) would of necessity be massively complex—since it will include, for instance, a specification of all the properties of all the particles in the universe—and there is no need to make it more complicated than it has to be.

3 A view something like this has been held by many theorists about qualia and experiences—perhaps even the majority, up until fairly recently—such as the Cartesian dualists, the psychological atomists (like Hume and Locke), the phenomenologists (such as Husserl and Heidegger), the logical atomists (e.g. Ayer, Wittgenstein and Russell), and the current crop of naturalistic panpsychists or “neutral monists” (e.g. Chalmers and Stubenberg). In using the word “atomic,” however, there are a variety of things connoted by the term in different contexts that I do not mean; instead of conjoining the content of all the various “atomistic” theories, I want to pick out a central aspect of the overlap of those theories. For example, in citing the logical atomism of Wittgenstein (1922) and Russell (e.g. 1918) I do not wish to necessarily commit my bland “atomist” to the claim that qualia-tokens are the basic units of meaning, let alone take a position on representational atomism (with, say, Fodor 1990, Dretske 1988 and Millikan 1984) as opposed to the holists (such as Harman 1982 and Field 1978). Similarly, I don’t intend to mean, necessarily, that qualia vocabulary has no atomic complexity (as in “atomic sentences”); and nor do I intend to make a link with the ancient philosophical doctrine of atomism (except, perhaps, insofar as both qualia-tokens and Democritus’ atoms are “impenetrable” to further analysis).
that it is a phenomenally red quale-token, but that no such further analysis or re-description is available through any method.\(^4\)

For an example, consider the difference between the properties inertial mass and transparency. Inertial mass is an “atomic” property in the current sense. It is defined as the tendency of an object to resist being moved (or if it is moving to resist a change in speed or direction) and this is all that can be said about this property even at the most complex level of description—this is what it is for even the smallest sub-atomic particle to have mass.\(^5\) By contrast, no sub-atomic

\(^4\) Note that one need not be understood as asserting that sensory experience is “atomic.” As explained in Chapter One, experiences—such as the visual image of a green tree on a lawn—are indeed analysable into component quælia-tokens, such as phenomenal greenness and brownness, ‘arranged’ in a certain way. Could the “atomist” say that some (though of course not all) quælia are analysable: are some quælia analysable into combinations of “basic” quælia? Is phenomenal orange, for example, analysable into some kind of combination of phenomenal red and phenomenal yellow? This is a rather difficult issue, but my inclination would be to discourage the atomist from taking this stand, for the following reason. We have established, I take it, that quælia are properties of states of the CNS and that the having of quælia is the experiencing of those quælia. For phenomenal orange to be a combination of phenomenal redness and phenomenal yellowness, then, it would have to be the case that some state of the CNS would have to have the quæla red, and some state the quæla yellow, and yet neither of these quælia would be experienced ‘as themselves,’ but rather they would somehow both together be phenomenal orange. This adds a significant amount of complexity and difficulty to one’s account of quælia. It is reasonably straightforward to understand, at least in outline, the claim that the quæla orange is a property of a state of the CNS—that to have an orange sensation is to be in a certain brain state. It is a great deal harder, I would have thought, to grasp what is meant by the claim that the experience of redness and the experience of yellowness can combine to be the experience of orange (when of course we are talking about phenomenal properties of the brain rather than pigments or lights). This is therefore, I suggest, a theoretical permutation to be avoided unless one has good reason to add it, and I am not aware of any good reasons for it. (The analogy with paint is, pretty obviously, a bad reason; and considerations to do with, say, the physiology of colour perception, such as opponent processing, have more relevance to how phenomenal brain states come about and in no way suggest that we phenomenally “code” only for the prime colours and black and white. Hardin (1988) and Van Gulick’s (1993) claims that quælia have structure probably tend to fall into this latter trap, but nevertheless deserve more careful consideration than space allows here.)

\(^5\) Actually, of course, since nothing is ever simple, a few caveats are in order. First, I am treating mass here as a non-relational property, one that is actually possessed intrinsically by objects. Second, I recognise that other true descriptions of the property are possible (such as weight divided by acceleration due to gravity), but assert that they just re-identify the very same property at the same level rather than providing an analysis of that property in terms of other more basic properties.
particle is transparent: transparency is explained in terms of a probabilistic summing of immense numbers of photon trajectories and exchanges (whereby electrons absorb and emit photons), combined with the effects of interference such as polarisation, and a material is transparent iff most of the photons hitting its surface have an amplitude such that they will be transmitted through the material. Transparency is thus not an atomic property, and it does not appear ‘as itself’ at the most complex level—all we need for a complete description of the universe, in the relevant respects, is information about the amplitudes of photons and electrons.

So qualia, then, are at a higher level of convenience only if (indeed, exactly if) they are not ontologically atomic. Claims about the level of convenience of qualia, therefore, have immediate metaphysical—and not just pragmatic or epistemological—consequences, even though these consequences follow merely from claims about the level of convenience of qualia talk, which is an epistemological-pragmatic concern. In particular, if qualia are not at a higher level of convenience then they are ontologically atomic properties and therefore (unlike most properties) feature ‘as themselves’ in any bottom level description or explanation or account of the universe. (We are not yet, however, talking about ‘levels of being’ properly speaking … though we are laying some of the groundwork for that discussion, below. As we shall see when we come to consider the metaphysical motivations for level-making, qualia could easily be at a higher level of convenience without being at a higher level of being; conversely, if qualia are ontologically atomic properties, but not what we shall call ‘basic properties,’ then they could be at a higher level of being despite being at the lowest level of convenience.)

So are qualia in fact atomic properties? The short answer, as usual, is we don’t yet know and it’s up to the empirical sciences to find out. Certainly, it is not inconceivable that they are not; they might turn out to be properties that are re-describable in more complex terms within a more detailed theory. For example, if Penrose and Hameroff are on the right track (e.g. Hameroff 1994, Penrose 1994), it could be that a TOE that links general relativity and quantum mechanics with a theory of quantum gravity will also provide a theory of qualia. Such a theory might show in detail how (and why) complex events involving quantum signalling along the microtubular structure of certain special neurones possess certain qualia. Note that the claim
here, if it is to be relevant, is not that such quantum effects bring about states which have qualia, but that the quale-token itself can be redescribed, reduced or analysed into a description in terms of, say, a quantum gravitational collapse of the superposed states of some critical mass of neuronal tubules. Under such circumstances it would turn out that qualia-terms exist at a higher level of convenience than TOE terms, just as predicates like “is water” and “is gold” capture equivalent descriptions which are far more complex.

Might Hameroff and Penrose, or other scientific reductionists, be on the right track? The evidence, we suggested in Chapter One, is not yet in. It does seem best at this point, however, to at least proceed as if qualia are not atomic: otherwise it is not at all clear that qualia will be amenable to study from a third-person perspective through the methods of science. The ultimate aim of a scientific theory of qualia is to predict and explain the fact that certain complex physical states of the CNS possess phenomenal properties. Typically, this would mean the theory shows that qualia are the phenomenal properties they are because of the way they are instanced in the brain; this is analogous with claims to the effect that, say, some piece of material \( A \) is transparent with such-and-such an index of refraction because of its quantum electrodynamical structure (i.e. its being transparent is nothing more than this structure). Thus, just as talk of transparency is really shorthand for talk in the language of quantum electrodynamics, so would qualia talk be shorthand for descriptions of complex physical properties of the brain.

On the other hand, if qualia are atomic properties then we would probably be faced with difficulties in studying them that are not encountered with other fundamental properties like mass or spin. This is because mass and spin are, of course, properties that are accessible through the “objective,” third-person methods of scientific observation; the “mode of access” for qualia, however, is primarily what is usually called “subjective”—it is the first-person experience of those qualia. That is, we apprehend the phenomenal property of redness by tokening that property in our brains and, as such, having the sensation of redness. If qualia are atomic then, presumably, there is little more to be said about the property itself—certainly, as we have seen, not anything like “and this quale is identical with some complex of third-person–identifiable properties.” It would seem to follow that qualia could be apprehended only subjec-
tively, and not via the usual scientific measurement methods. This is perhaps what Owen Flanagan means by the hypothesis (which he rejects) that “the properties of qualia are objectively undetectable because the properties of qualia are exhausted by their subjectively available aspects” (1992, 69).

Those whom I labelled “reductive physicalists” in Chapter One, then, must hold that qualia are non-atomic, and thus at a higher level of convenience than the ‘bottom level’; by contrast, “hard liners” should probably assume that qualia are atomic, and that qualia talk is, roughly, a maximally detailed and explanatory way of denoting phenomenal properties. As to which view is correct, the jury is still out … but those of a physicalist bent who are genuinely interested in explaining qualia must take their non-atomicity as a working assumption until shown otherwise.

b) Modelling and Theory-Relativity

Let us move on to the use of levels in modelling. A model, in science, is, generally, “a theory intended to explain a given realm of phenomena” (Lacey 1996, 219), and in a more specific sense is a representation of one system—that being studied—by another whose workings are supposed to be analogous in the relevant respects to that of the first. This second, analogous, system may be either an already understood piece of reality (such as water waves as a model for sound propagation) or a theoretical version of the first system with various simplifying assumptions (such as, for example, Crick and Watson’s original double-helix model of DNA, or the gene model of inheritance). Ronald Giere calls the first sort an “analogue model” and the second a “theoretical model” (1991, 23–28).

Qualia would normally only be viewed as being at the relatively abstract level of a model if they essentially form part of some model; since it is fairly obvious that qualia are not features of an analogue model (they are not an already understood system representing some other sys-

---

6 The only way around this ‘problem’ with atomic qualia that I can see is the possibility that qualia could be identical with some one, objectively observable, fundamental property … which, presumably, has yet to be discovered, since none of the extant atomic physical properties (such as mass and charge) are correlated exactly with qualia.
tem), this will be the case only if qualia are terms that find their meaning within a theory and its
simplifying assumptions. If qualia have a status analogous to the force of gravity in Newton's
physics, then they are candidates for being terms of a higher-level model, rather than lower
level items; if they have a status closer to that of the perturbations of the perihelion of Mer-
cury or the accelerating speed with which apples fall from the Tower of Pisa, then they do not
belong at the level of a model, but rather at the level of that which the models might try to
explain.

So are qualia relative to a model? Why might one think so? One standard way of trying to
make the case that they are model-relative is by asserting that the concept “qualia” finds its
meaning within a web of theory about, for example, first-person incorrigibility, ineffability or
even incorporeality. That is, the claim is that qualia are whatever it is that fits that conceptual
space—whatever it is that is incorrigible, ineffable and incorporeal, for example—and if noth-
ing fits that space, then the term “qualia” has no actual referent. Similarly, if there were no
force describable (sufficiently closely) by Newton’s equations, then gravity—or at least Newto-
nian gravity—would not exist (though of course falling apples still would, and we would have
to find some other explanation for them).

However, the notion of qualia used in this dissertation does not get its meaning in this way,
and so is not dependent upon a model like this. We began by announcing that qualia are simply
the colours, sounds, smells, etc. that we experience. When one sees a red wall, that colour
is the quale redness, and it is that colour for which we require a theory. Whether that phenomenal
redness is identical with any actual property of the wall, whether our apprehension of it is in-
corrigible or not, whether qualia are “physical” or “mental” … none of these things are already
part of the concept quale, but are instead further things to be discovered about qualia before

---

1 Indeed, those who make this kind of claim about qualia—or phenomenal consciousness generally—typically do
so with the aim of showing that qualia or consciousness do not exist. The work of Gilbert Ryle (1949), Patricia
Churchland (1983), Georges Rey (1983) and Daniel Dennett (1988) shows distinct signs of this tendency. Stephen
Stich’s book Deconstructing the Mind (1997) argues that most important forms of eliminative materialism in the phi-
losophy of mind have all relied upon theory-relativism.
we can give a full account of them. To approach the same point from another angle, the notion of qualia does not explain anything, as the notion of gravity does; rather, it picks out something which is to be explained.⁸

One response to this claim, however, is to deny that there is any force to the observation/theory distinction upon which it relies.⁹ From this one might argue that, since all terms are to some degree or another theory-laden, so too is the term “qualia.” Thus, qualia observations are not ‘pure’ observations but take place through the lens of theory; change the theory, and so too will the observations change. And indeed, this is one way of characterising Paul Churchland’s well-known thesis of “eliminative materialism.” His claim (e.g. 1988a, 47–49) is not directly that mental states like qualia (strictly speaking, experiences) are theoretical constructs but that qualia observations are theoretically loaded: to make claims about qualia “makes the same mistake that an ancient or medieval person would be making if he insisted that he could just see with his own eyes that the heavens form a turning sphere, or that witches exist” (1988a, 47). This is what Churchland has called the “plasticity” of all the ‘forms of perception,’ including introspection (1979, 7 ff.; 1988b).

What shall we say about this version of the theory-relativity claim? First, suppose we accept that all observations are to some degree theory-laden; it does not follow from this that qualia are more like Newtonian gravity than like falling apples or orbital perturbations. What is at issue, one might say, is which body of ‘theory’ informs and gives meaning to the observations. Some ‘theories’ are, as we might put it, distinctly more ‘specialised,’ more liable to change, and even more dispensable than others, and the degree of theory-ladenness involved is correspondingly significantly different. The change from Newtonian physics to Relativity theory, for example, brought about changes in various terms and observations (such as those involving

---

⁸ John Searle and Owen Flanagan similarly insist that qualia are important elemental explananda for any theory of mentality, and hence not just the terms of some model. See, for instance, Searle 1992, especially chapter 1, and Flanagan 1992, 21 ff..

⁹ Influential early detractors of this distinction were, for example, N.R. Hanson (e.g. 1961) and Paul Feyerabend (e.g. 1988).
gravity), but left unaffected most of our observation-statements about apples; presumably it would require at least a fairly radical change in the English language to have that kind of effect, such that for example the noun “apple” no longer exists or changes its meaning. Even more radically, as Churchland himself admits (1988b, 140 ff.), some of the relevant ‘theory’—assumptions about the three-dimensionality of space, the spatial and temporal continuity of common objects, the sharp change of luminance at a body’s boundaries, colour constancy through changing environments, the occlusion of distant bodies by proximate ones, etc.—could possibly be hard-wired into the normal Homo sapiens brain, and so would de facto remain constant across most human beings (whatever their language).

This is not, it seems to me, a trivial point. It suggests that we can hold open at least a rough and ready distinction, at least towards the ends of the spectrum, between being relative to a scientific theory and being relative to other bodies of ‘theory’ (like the assumptions embedded in our brains, natural languages and forms of life). And, for our present purposes—dealing with models—it is really explicitly scientific theories that we are concerned with; the English language is hardly a “theoretical model” in the sense that Giere uses the phrase, for example. Where do qualia fall along this spectrum of theory-ladenness? It seems most likely that they are far closer to the physiology-laden end of the continuum than they are to being relative to a theoretical model within one of the sciences.

Second, we should note that this argument about theory-relativity applies to judgements about qualia, and not to qualia themselves. Churchland, for example, intends his points to apply to “introspective observations,” where introspection is thought of as “the disposition-governed occurrence of conceptual responses to one’s internal states, responses made within whatever matrix of self-understanding one has developed or acquired. Our current matrix of response is

---

10 “Who ever claimed that the character of a scientist’s perception is changed simply and directly by his embracing a novel belief?” (Paul Churchland 1988b, 146).

31 I make a more detailed argument to this effect, with an aim to discovering a widely-intersubjective basis for the comparison and assessment of scientific theories, in my paper “Split-Level Contexts, Intersubjective Facts and Semi-Naïve Empiricism: A Semi-Naïve Fairy Story” (1997b). I have since found that Jerry Fodor has argued along parallel lines for the “rigidity” of perception (Fodor 1984).
the P-theory [roughly, folk psychology]…” (1979, 116). He holds it to be “just conceivable” that “all of our introspective judgements have been systematically false by reason of presupposing a false background theory” (1979, 96), and thus that it might be that there are really no such properties as qualia.\(^\text{12}\)

Where I differ most centrally from Churchland’s account of introspection is in holding that some of these “internal states” are *phenomenal*, and are therefore *already apprehended*, independently of the “conceptual responses” they give rise to. Qualia are the colours, sounds, smells etc. of our sensory experience; these are phenomenal properties—to be in a state possessing these properties just is to have a sensation of the relevant sort. Could we, then, replace qualia with introspective judgements of “glucose consumption in the forebrain, dopamine levels in the thalamus, the coding vectors in specific neural pathways, resonances in the \(n\text{th}\) layer of the peristriatal cortex, and countless other neurophysiological and neurofunctional niceties” (Churchland 1988a, 180)? No—or at least not in the wholesale, first-person way Churchland envisages. It is possible, certainly, that we could discover identity relations between the colours we experience and, say, coding vectors; we could then say to ourselves, “That shade of phenomenal red that I am currently experiencing is neural property \(N\) of brain area \(s\)—that is another way of describing what I am experiencing.” Given an adequately explanatory theory, this neural way of talking might be as good or even better than the qualia vocabulary. But it is not the case that the phenomenal redness would somehow *disappear* once we started talking in this new way, or that we would experience property \(N\) ‘directly’ in a *new* way, perhaps closely analogous to the third-person access to \(N\): by contrast, phenomenal red would still be the first-person, and \(N\) the third-person, way of apprehending the *same* property.\(^\text{13}\)

---

\(^{12}\) “My own 1979 position … simply assumes the generally constant character of our sensory responses to the environment. The plasticity that excited me there was confined to the conceptual frameworks within which we make our judgmental responses to the passing contents of our sensory manifold” (Paul Churchland 1988b, 151).

\(^{13}\) Nevertheless, one could (just about) imagine some psychological disorder by which the more impressionable among us cease to believe in their own qualia. (I am assuming these people are still physiologically sufficiently normal that they do *have* qualia; they have not actually been zombified by their disease, only their cognitive states have been affected.) Sufferers would presumably sincerely assert things like, “I do *not* experience any shade of
My conclusion, therefore, must be that qualia do not form part of any levels-structure constructed for modelling purposes, and that this is in part because they are not in any interesting way theoretical, model-relative, constructs.

c) Explanatory Levels and Epiphenomenalism

What about the third variety of pragmatic or epistemological levels in the taxonomy of the previous chapter, explanatory levels? This kind of invocation of hierarchies of levels, as we saw, arises from the nature of explanation itself: in order to construct good explanations we are often required to operate at higher levels. This variety of levels, though as we have seen not identical with hierarchies of levels of convenience, does entail the existence of levels of convenience: some description or explanation could not be at a higher level of explanation if it were not also at a higher level of convenience, for otherwise there would be no simpler explanation available (at least not one simpler in virtue of being at a lower level of detail). So qualia will not be at a higher level of explanation if they are not also at a higher level of convenience. We arrived at the tentative conclusion above (section a)) that qualia are indeed at a higher level of convenience, and interestingly considerations to do with levels of explanation seem to place us in a dilemma which reinforces this conclusion.

Suppose, first, that qualia are not at the lowest level of convenience. Then, it seems fairly clear, we do often invoke a higher level than the lowest in order to use qualia-terms in explanation. For example, one good explanation in a certain context for why I snatched back my hand from the candle flame might be that I felt a burning pain sensation and reacted to it. As we saw phenomenal red; all that I experience is the \( N \)-ness of my own brain state \( s \), and this experience is not phenomenally red but (perhaps) rather like what you experience when you detect the fact that my \( s \) is \( N \).” This disorder is possible only because of the distinction between qualia and beliefs about qualia: as Churchland points out, it is possible for our conceptual responses to qualia to be systematically wrong. However, this disease would be a very peculiar one, much more so than Churchland envisages; because there would be phenomenal redness going on, and therefore the sufferer would, of necessity, phenomenally apprehend that redness; they just would not believe in it. This would be not so much like seeing a ghost without believing it is actual, as like seeing a ghost but not believing—or noticing—that one is having that sort of sensation … that there is even apparently an apparition.
in section 2c of Chapter Nine, it would not be relevant to this explanation to describe the compound nature of this pain (how it was analysable into physical states $x_1 - x_n$, for example), and replacing the term “a pain” with one of these descriptions would make the explanation *markedly* less satisfactory in very many contexts. In other words, if qualia terms pick out ‘higher-level’ entities then those terms are necessary in order to make certain standard kinds of good explanation.

Suppose, on the other hand (*for reductio*), that qualia are at the lowest level of convenience. This would entail, we said in section a) above, that qualia are ontologically atomic properties, and if qualia are atomic then we can quickly show that this combines with standard assumptions to suggest that qualia are *epiphenomenal*—that is, that they would not form part of *any* accurate explanations (let alone good explanations) at that level at all. If human behaviour can be completely explained by the operations of the brain, which in turn can in principle be completely explained in terms of ‘objective’ atomic properties, then though atomic qualia might exist as ‘ontological danglers’ they could not be required at the bottom level for complete explanation. However, in very many contexts, qualia are required for good (or even adequate) explanation. If a father wants to know why his baby is crying, he will not be satisfied with—perhaps will not even understand—an explanation in terms of afferent signals, brain processes, and efferent signals. He will want to know if the baby *feels* a sensation of wetness, or is in pain, or is feeling tired. These are qualia terms, and so qualia cannot be atomic: if qualia talk were at the ‘bottom’ level of explanation and description then it would not be explanatory at all; qualia talk is explanatory; therefore, it must be at a higher level of explanation (and thus description) than the lowest.

Though this is in my view a somewhat weak argument (mainly in virtue of taking ‘standard assumptions’ rather uncritically), I take it to be suggestive: added to the considerations to do with qualia atomicity adduced in section a) above, I shall tentatively conclude that qualia are in fact at a higher than basic level of explanation.
2. **Ontological Levels and Qualia**

We have said that qualia might (or might not) be re-describable in more complex ways at a lower level,\(^{14}\) and that qualia terms might disappear at that lower level: we have not yet directly addressed the question of what this might mean. I suggested in the last chapter that there are two identifiably different kinds of ontological levels: mereological levels and levels of being. Here, as in the previous section, I want to apply this framework to qualia and see what shapes emerge.

\(\text{a) Mereological Levels, Basic Properties and Emergence}\)

One reason, then, why there can be an ontological need to ascend to higher levels is the fact that certain groups, sets or wholes can have properties that are not properties of any of their members or components. Indeed, we have seen, \(\text{most}\) properties are like this—they are not also properties of the fundamental elements that make up the objects in which they inhere. Only the properties of the most fundamental constituents of the universe—I shall call these “basic properties”—are not higher-level properties in this sense. A few properties are simultaneously basic and higher level: charge, for example, is a property of the fundamental elements of the universe (a quark has a charge of \(-\frac{1}{3}\) or \(+\frac{2}{3}\) the charge of an electron, for instance) yet it is also a higher-level property belonging to individuals ranging from hadrons to electrical substation to storm-clouds. True, the charge of a battery is presumably somehow summed in a very complex way from the charges of all the elemental particles making it up, but one wants nevertheless to treat it as the same property.\(^{15}\)

So are qualia basic properties or what I have called group-level properties? And if the former, are they properties only of the fundamental elements of an ultimate Theory of Everything (TOE), or are they also higher-level properties? We can deal with the second of these ques-

\(^{14}\) And \(\text{probably}\), therefore, be explicable by that lower level.

\(^{15}\) Just as we would consider “pinkness” a basic property if all the elemental components of a pink ice cube were pink.
tions rather quickly, given the theoretical framework we have been constructing about qualia: since, we have said, qualia terms pick out properties of CNS states, and since CNS states are not fundamental entities of the TOE (but are higher-level individuals), then qualia cannot be properties of only basic elements. The first question on the other hand—are they, then, basic properties at all?—is less straightforward.

Considerations about the atomicity of qualia (see section 1a)) are relevant here: in short, if qualia are atomic then they are probably also basic; if they are not atomic then they are probably also not basic. Suppose that qualia are not ontologically atomic (as I have suggested we should assume they are not); it follows from this, by definition, that qualia are analysable into collocations of lower-level properties (as transparency is reducible to quantum electrodynamical properties), and these properties, it seems reasonable to suggest, will be properties of more basic individuals. Just as transparency is not a property of any of the particles that make up a transparent object, neither would qualia be properties of the component elements of a qualifer such as a phenomenally red CNS state.

Suppose, on the other hand, that qualia were atomic. This, by itself, tells us nothing about what these atomic properties are properties of—that is, that a property is atomic does not actually entail it is, or is not, basic. However, we can infer from the claim that qualia were both atomic and not basic that qualia would have to be at a higher “level of being”\textsuperscript{16}: thus if we can show (as I shall try to in the next section) that qualia are probably not at a higher level of being, then if qualia turn out to be atomic after all, it would follow fairly directly that they are very likely to be basic properties.\textsuperscript{17}

Our provisional conclusion then, is that qualia probably would be basic properties if they were atomic, but that since we have held it to be more likely that qualia are not atomic we

\textsuperscript{16} Since qualia are ineliminable and would not correspond to any lower level patternings: this fulfils our definition of “being at a higher level of being than” in section 3b of Chapter Nine. (See the next section for more on this.) Note that if qualia do turn out to be basic properties then it will also follow that they cannot be at a higher level of being since they will form patterns of similarity at the bottom level.

\textsuperscript{17} That is, if qualia are not at a higher level of being they must be either non-atomic or basic; if they are atomic, then, they must also be basic.
Chapter 10: “Spirit Levels.”

should also assume they are not basic.

One final consideration might be thought to count somewhat against the hypothesis that qualia are basic properties (and thus also that they are not atomic): this is the suggestion that treating qualia as basic properties sits rather uncomfortably with the not infrequently held opinion that they are ‘emergent’ properties.\(^{18}\) So, even if qualia are basic properties, we can still ask whether they are emergent properties or not, in the following sense: Are qualia properties of every one of some particular class of basic elements, no matter how they are situated, or are they properties that “come into existence” only when these basic elements are combined in a certain way—even though they are still properties of those basic elements? This notion can be expressed either synchronically or diachronically, but has a greater force when considered synchronically. The sort of situation I imagine is one where there is a kind of fundamental element \(E\) and a basic property \(P\). At spatiotemporal region \(r_1\), element \(e_1\) (of type \(E\)) lacks this property \(P\); however element \(e_2\) (also of type \(E\)) at region \(r_2\) possesses property \(P\) (even though it is the same kind of fundamental element). By hypothesis, \(e_2\) possesses \(P\) while \(e_1\) lacks it because \(e_2\) is located in some kind of complex combination of fundamental elements, upon which property \(P\) is emergent, even though it is still a basic property: that is, bottom-level property \(P\) is present only in those spatio-temporal regions with, say, a particular complex structure.\(^{19}\)

Now, one reasonable response to this putative sub-type of emergence is to object that it is incoherent: that it makes no sense, at the bottom level, to say that two fundamental particles, one of which has intrinsic property \(P\) and the other of which does not, belong to the same type \((E)\). That is, it is plausible to assert that fundamental elements are categorised according to all their properties.\(^{20}\) It follows from this objection that the assertion that some property \(P\) is both

---

\(^{18}\) As expressed by the British Emergentists, for example: see Horgan 1993 for a useful summary here, building upon McLaughlin 1992.

\(^{19}\) The contrasting, non-emergent, case, of course, would just be one where every fundamental individual of kind \(E\) would (in every region and at every time) have basic property \(P\). For example, we might simply say that electrons have a negative unit charge—no emergence is involved here.

\(^{20}\) A “top quark” of a certain “colour,” for example is a type of fundamental particle with a particular mass,
emergent and basic entails that the emergence of this property changes some of the fundamental elements which compose the instancing object. For example, it follows from the British Emergentists’ claim that phenomenal mentality is a “fundamental” property which is emergent from sufficiently complex physico-chemical concatenations (e.g. human brains), that some of the elemental components of human brains were replaced by elements of a different type at the moment of emergence (and presumably that this is so because of the increased complexity of the whole). Thus if qualia were basic properties then this would render the claim that qualia are emergent an even more interesting one than is usual … and correspondingly less plausible prima facie. This in turn might furnish some additional reason to think that qualia are in fact not basic.

b) Levels of Being and Irreducibility

Let’s assume, then, that qualia are at an ontologically higher level in virtue of not being basic properties: this is really not a terrifically interesting or surprising metaphysical conclusion—virtually all properties, many of them perfectly respectable scientific structural and functional properties, are at a higher mereological level. What would be much more interesting would be if qualia were at a higher “level of being” than, say, the neuronal level—if they were metaphysically in a completely different ‘compartment’ than the physical. I suggested in the previous chapter that this should be taken to mean that qualia are both ontologically (as opposed to epistemologically and so on) ineliminable and irreducible to fundamental physical properties. I shall argue here that, though this a matter ultimately to be decided a posteriori, we have some reason to think qualia (at least assuming they are not atomic) are not so irreducible. This is just charge and spin: any particle that comes along with a different mass, even if all its other properties were like those of a red top quark, would be a different sort of particle.

21 That is, if by this they meant that it is a basic, rather than just an atomic, property.

22 In fact, one might say, it would be more surprising if qualia were basic properties, like charm and spin—if quarks were phenomenally red.

23 Or “the rest of the physical”—see below for a discussion of this point.
because, I shall argue, it would be very odd for any property to be ontologically irreducible to properties at a lower-level on which it does not itself appear.

Without wishing to engage with the important and difficult debate about what properties exactly are, I assume that at a minimum properties are a subset of the complete set of what I have been calling (Chapter Nine, section 3b) “patterns of similarity.” I am treating the phrase “pattern of similarity” extremely broadly: I think, in fact, that patterns of similarity can be understood as underlying kinds of conditions for set membership—actually, any condition other than simply membership in the set itself. Further, I take patterns of similarity to be robust ontological entities and not, for example, linguistic or logical entities; when I describe them as “conditions for set membership” I mean that set members are all similar in a certain—perhaps very complex—way, and that no non-members of that set are similar in that way. Nor is a pattern of similarity to be treated as being ‘projected,’ or relative to a certain viewpoint, or in some other way irreal. Either a group of entities resemble each other in some way, or they do not, and that is an objective fact of the matter about the world. Some obscure pattern of similarity—such as being both light blue and less than four feet wide—is no less “real” in this sense than a “genuine” property like having a mass of one kilogram: certain objects really do resemble each other in being light blue and less than four feet wide.24

As outlined above, I take a higher level property, such as a quale, to be ontologically irreducible iff it corresponds to no bottom level pattern, since that is exactly the claim that there is not even potentially a low-level property to be picked out with which it is identical. Thus, qualia are irreducible in the sense we are interested in here iff it could be the case that either:

i) two or more realisations of quale $Q$ have no low-level properties in common, or

ii) two low-level tokens could be qualitatively identical (at that level) but one be an instantia-

24 Let me note two caveats here. Firstly, the universe is a fuzzy place and contains degrees of similarity: this indeterminacy is perfectly ‘objective,’ and does not make patterns of similarity any less ‘real.’ Secondly, certain kinds of similarity, though in a sense objective, might be in certain ways relative or private; thus, I might group the objects in a room according to whether I find them aesthetically attractive or not, but as long as there is a fact of the matter about my tastes at the time, then there is also a fact of the matter about the pattern—that a description of the pattern must make reference to my tastes does not itself make the pattern irreal.
tion of $Q$ and the other not.

Under these conditions, even if it were possible to form a ‘predicate’ by disjoining labels for all of the realisations of $Q$, that ‘predicate’ would still not pick out even a real pattern, let alone a property proper. On the other hand, under any other conditions, the tokenings of quale $Q$ would also be tokenings of a low-level pattern … which would probably be in itself good enough reason to treat this pattern as a *bona fide* low-level property ($P$), to which $Q$ is reducible.

I doubt that it is possible from the armchair to *prove* that qualia are ontologically reducible, because I believe this is an empirical question which has yet to be decided (and which probably won’t be decided for several years to come). Yet it seems to me that one can make the following informal argument to suggest that qualia are *prima facie* unlikely to be at a higher level of being.

Suppose it were the case that two physical individuals were qualitatively identical in terms of their low-level properties, including any relevant relational ones such as being part of some especially complex system, and excluding only the relational properties that necessarily uniquely identify an individual (such as exactly occupying spatio-temporal location $x$, or being identical with $i$). And suppose that one of these two individuals has the higher-level phenomenal property $Q$ but that the other either instances no quale at all or some other quale, $R$. That is, suppose condition ii) above is fulfilled. It would follow directly that qualia do not after all supervene in *any way* upon low-level properties: they would not be related to basic properties (or states, etc., described in terms of those properties) by even the minimalist relation of Global Token Supervenience, which we agreed at the start of Chapter Eight was tacitly accepted by almost everyone who thinks about qualia (including those who like to think of them as being interestingly ‘higher level’). It would mean that fixing the bottom level of basic properties would do *nothing whatsoever* to fix qualia even merely in the actual world, never mind in the possible worlds next door.

This consequence would make qualia a great deal more ontologically strange than any of the other run-of-the-mill examples of ‘irreducible’ properties, such as moral, aesthetic or economic properties. To put it another way, moral, aesthetic and economic properties, though ‘irreduci-
ble’ and ‘higher level’ in at least some of the ways outlined in Chapter Nine, are not properties which occupy a higher level of being; if qualia did so, they would be alone in this exalted and mysterious status.

Consider, for example, the dramatic beauty of Lawren Harris’ painting *Decorative Landscape*. Suppose one held constant all of the low-level physical properties of all the relevant individuals involved: the painting itself, ‘expert’ beliefs about and attitudes towards the painting, the states of mind of Harris as he was painting it, the circumstances of its being painted, and so on. Suppose also, so as not to beg any questions, that we hold constant all the qualia involved in perceptions of the painting. And finally suppose that the low-level physical properties fix the higher-level properties that make the aforementioned individuals the things they are—say, a belief that \( p \), or a painting of a yellow sky. Now, it is usual to assume that having made all these suppositions we are no longer at liberty to vary the aesthetic value of the painting … the colours, shapes and composition of the picture, perhaps combined with its value in the eyes of art experts, its status as an expression by the artist, and suchlike, just constitute the painting’s beauty; they exhaust the reasons why it is beautiful. Aesthetic properties supervene upon the relevant set of low-level properties. To deny this would be to say something quite odd about beauty; it would be to suggest that some picture \( a \) can be beautiful but another, \( b \), which looks exactly the same, has the same importance in the history of art, is equally authentic, is equally well regarded by sensitive observers, and so on, is not beautiful. No one thinks this about beauty; however this is the kind of claim that this non-reducibility thesis about qualia (in one of its disjunctive guises) forces upon us.

Similar oddities follow from the other disjunct above: the claim that two individuals that token some quale type might have no low-level properties in common. Consider, for example, being a murder, being worth one Canadian dollar or being a letter-opener: all of these are quite central examples of multiply realizable, ‘high-level,’ pragmatically irreducible properties. Suppose that two events \( a \) and \( b \) are both murders; suppose also that the cessation of any electrical activity in the brain corresponds to some low-level complex property or set of properties, which I shall call \( D \). Let us say that \( a \) involves some individual that has property \( D \)—that is,
roughly, the murder involves someone’s death. Now, the claim we are testing entails that event \( b \) is utterly different from \( a \) at the low level in question, which means that \( b \) involves no individual with property \( D \) … which would mean that no one dies as a result of event \( b \), which means that \( b \) cannot be a murder. Yet \( b \) is a murder, so in the case of at least this high-level property, the assumption that two individuals that possess it can have no low-level property at all in common is absurd.

Parallel reasoning goes through for the other two examples. Suppose item \( a \) has the property that someone would be disposed, under some set of circumstances, to swap a genuine Canadian ‘loony’ for that item. This disposition is a physical property of an individual and, presumably, is describable as a low-level property \( V \). Suppose, *ex hypothesi*, that \( b \) does not possess \( V \); then no one is disposed to pay $1 for \( b \); so \( b \) cannot be worth $1. Or suppose \( a \) has the property that it is sufficiently rigid to tear paper, yet \( b \) lacks the corresponding low-level property—then \( b \) cannot possibly be a letter opener. And so on. By contrast, the claim with respect to qualia is precisely that, for any low-level property of quale-tokening \( a \) you care to pick out, there is some quale-tokening \( b \) that lacks it: since it is absurd for other common examples of high-level properties, it is at least peculiar for qualia.

I conclude provisionally, therefore, that qualia are unlikely to be ontologically irreducible in the sense that they exist at a higher level of being. However this argument would be no match for hard empirical results; if scientific investigation established that qualia correspond with no low-level pattern, then we would be forced to accept one or both of the disjuncts i) or ii) rather than deny those experimental results. It would not be unique in the history of science that experimental results falsify plausible *a priori* intuitions.

Imagine for a moment that a day comes when qualia *are* found to be ontologically irreducible. The pressing question then, it might seem—indeed, the question that many find pressing today—will be “Are qualia uniquely *mental* (or anyway non-physical) properties?” Yet it seems to me that the really important metaphysical question will have *already* been answered by the time we get to this stage: on that day it will already be known that qualia are ontologically irreducible properties of physical individuals. What remains is merely more or less a matter of la-
belling convention—shall we call such properties “physical” or “mental”? The results of the previous chapters give us no reason to believe that qualia are somehow ectoplasmic or unworldly—they are, we have found, just properties of states of the CNS, albeit perhaps bizarre ones. Indeed, we should probably have to insist that, were we to remove everything physical from the universe, *nothing* would remain, not even the space-time continuum, and in particular not disembodied qualia (if only because there are no unowned property tokens). Why, then, should we call qualia “mental,” except perhaps just in honour of their (supposed) ontological irreducibility; but if *that’s* what we are doing then to say qualia are specially “mental properties” adds nothing interesting or substantive whatsoever to what we have (*ex hypothesi*) already found out about the metaphysics of qualia.

Qualia, then, are probably not ontologically irreducible properties—they probably do not occupy any higher level of being—and even if they were it is not clear that this would mean they are “non-physical” in any interesting sense *beyond* that of being ontologically irreducible.

3. CONCLUSION

So are qualia, as is often said, at a “higher level” than (though dependent upon) ‘the brain’? Are they somehow invisible or even non-existent at the ‘level’ of bosons and fermions, or that of synapses and neurotransmitters? It turns out, of course, that it depends on what you mean by “at a higher level than,” but that we have no good reason to suppose qualia are ‘higher level’ in any particularly novel or difficult way. I have argued that we should treat qualia as being non-atomic properties—that is, as properties which are analysable into other, more basic, properties—and that they are therefore somewhere above the ‘bottom’ level of convenience; this conclusion, supported by the fact that qualia do in fact feature in certain kinds of good explanation, shows that qualia are probably also at a higher level of explanation. On the other hand,

---

25 One can imagine a future neuroscientist saying to herself, “Now! We have finally discovered that qualia are irreducible higher-level properties of chunks of the CNS; let’s pull up our sleeves and do some research to find out more about these fascinating properties.” It seems to me unlikely that the questions that would strike her as the most fruitful or interesting would include “Are qualia physical?”
I have tried to show that qualia are not merely constructs in some theory about the brain, and so do not occupy a modelling level.

Ontologically, qualia are most likely at a higher mereological level—they are not basic properties: properties of the elemental entities of a completed physics—since we have suggested that they are non-atomic properties. However, they are much less likely to exist at a higher level of being than the rest of the universe: at a minimum, such a claim must be recognised as the radical one it is, implying as it does that, for example, qualia violate supervenience. We are left, then, with no good reasons to suppose that qualia are ‘higher level’ in any more interesting or radical metaphysical sense than are can openers and aesthetic properties.
Chapter 11: Three Central Counter-Theories

I have now completed my positive account of the metaphysics and epistemology of qualia. In this penultimate chapter I will briefly consider, and show why I reject, three accounts of qualia that have been influential but which are directly opposed to the one I have defended and elaborated over the last ten chapters. Each of the three represents a significant strand in the different modern approaches to qualia: Daniel Dennett represents the eliminativists, Fred Dretske the externalists, and David Chalmers the property dualists.

1. Dennett’s Eliminativism

Daniel Dennett is the best-known of the qualia eliminativists. His 1988 article “Quining Qualia” is arguably the central recent paper on the topic, and certainly already one of the most anthologised. I shall here consider that paper in some detail. It contains, I will suggest, two inter-twined but distinguishable strands of argument, one more explicit than the other, but both faulty.¹

a) The Four Properties

Though it is ultimately a rather subtle and elusive paper, Dennett presents “Quining Qualia” as if he is making just the following fairly straightforward point: he announces that, although he is willing to talk of the properties of conscious experience, these properties

...are so unlike the properties traditionally imputed to consciousness that it would be grossly misleading to call any of them the long-sought qualia. Qualia are supposed to be special properties, in some hard-to-define way. My claim ... is that conscious experience has no properties that are special in any of the ways qualia have been supposed to

¹ Dennett’s qualia eliminativism is also expressed and argued for, along similar lines though with his crypto-verificationist inclinations often more evident, in his 1978a, 1981, 1991b, 1992 (with Marcel Kinsbourne) and 1994. He recently announced, however, at a session of the 1998 Central Division meeting of the American Philosophical Association, that he is no longer an active advocate of qualia eliminativism and considers the battle for qualia eliminativism to have been lost.
That is, none of the “properties of consciousness” count as qualia because they do not have
the second-order properties which Dennett thinks of as definitive of qualia (46–47), namely:
1) Ineffability;
2) Intrinsicality;
3) Privacy;
4) Being the objects of direct or immediate apprehension in consciousness.

This kind of claim is in a direct line of descent from those of Ryle, the Churchlands and
Georges Rey, considered in the section on model levels above (Chapter Ten, section 1b), and
part of my response to this kind of argument has already given there. In brief, an argument of
this sort does not refute the existence of qualia, or even show there is nothing we can say
about them; at best it refutes a particular theory of qualia. The question is, then, what is the
status of that beleaguered theory?

Most probably, the theory that Dennett attacks is stuffed with straw. As Owen Flanagan has
said, “[t]o be sure, all the properties he lists have been ascribed by someone or other to qualia.
But I can’t think of one credible recent philosophical source that uses the term in a way that
ascribes all or even most of the features he lists” (1992, 73). Dennett does predict such a criti-
cism and in response rejects the claim that there is a “much blander and hence less vulnerable
notion of qualia” (47) available, in effect arguing that if there is any traditional notion of qualia
at all then it is his four-fold characterisation. Yet Dennett earlier in the paper uses just such a
bland notion himself to pick out for the reader what it is he is bent upon ‘quining’ “‘Qualia’ is
an unfamiliar term for something that could not be more familiar to each of us: the ways things
seem to us” (42), he writes, and he gives as his examples “the particular, personal, subjective vis-
ual quality” of a glass of milk at sunset, as well as the “way the milk tastes to you then … and how it

2 Although this is the way Dennett sets out his argument in “Quining Qualia,” it might be more natural to inter-
pret him as ‘really meaning’ that the experiencing of qualia is (putatively) a part or kind of consciousness—or
possibly of conscious content—but that no part/kind of consciousness has the four second-order properties and so
none is the experiencing of qualia.
sounds to you as you swallow. ... [L]et the entire universe be some vast illusion, some mere fig-
ment of Descartes’ evil demon, and yet what the figment is made of (for you) will be the qualia of
your hallucinatory experiences” (42). And this, or something like it, is “the source concept,
the ‘pre-theoretical’ notion of which ... [technical concepts] are presumed to be refinements”
which Dennett professes himself concerned to debunk (44) but does not address. Qualia are,
minimally, just the colours, sounds, smells, (pains, emotions, mental verbalising?) etc. that con-
stitute our conscious experience, whether or not they are private, ineffable, intrinsic and immedi-
ate.

Dennett’s response to this criticism can only be that the bland notion he himself uses is
merely shorthand for this richer four-property conception (which I call the “pi-3” theory); that,
really, when we talk about the colour, taste and smell of milk (in this context) we have in mind
the pi-3 theory. All I can say, as I have pointed out above, is that at any rate I do not ‘really
mean’ the term in this way. When I say that the taste of milk is a quale, then that is just what I
mean and nothing more—that property, whatever it is, is an instance of what we have been
concerned with for the past two hundred and fifty pages or so. Is that property internal or ex-
ternal—is it identical with some property of an external object (the milk)? Are first-person be-
liefs about it incorrigible or indubitable? In what sense, if any, is the property itself or our ap-
prehension of that property “private”? All these questions arise after we have identified what
property it is we are talking about, they are not part of the process of that identification.

Whether or not Dennett’s attack actually does address the canonical notion of qualia, how-
ever—whether it impacts upon “the very meaning of the term”—there is still a sense in which
we should still consider its force against my theory of qualia, as I have elaborated it here.
Though I do not make Dennett’s four properties constitutive of the very meaning of the term

---

3 It’s true that Dennett is presumably just setting the stage by defining what is typically meant by “qualia” here,
rather than actually endorsing these claims. But that’s just the point: this is what is typically meant by “qualia.”
4 He describes himself as arguing that all those who speak of qualia must hold the pi-3 theory, and furthermore
that all those endorsing pi-3 are “dualists or mysterians or anti-functionalists of one sort or another” (oral pres-
entation at the 1998 Central Division APA meeting). I dispute both these claims here.
“ quale,” I do in fact hold that qualia are, in some sense, intrinsic, private and immediately apprehensible (in the ways now made clear in my chapters discussing these properties, especially Chapters Four and Six) and, though I have not addressed the matter of their ineffability I am prepared to admit that there is a sense in which they are ineffable as well. I therefore appear on the face of things to be a good target for Dennett’s deflationary zeal after all. Nevertheless, my view of qualia is in fact a great distance from that which Dennett seems to attribute to the qualophile and, in my view, a great deal more reasonable and defensible. I hold that qualia might be called “private,” “immediate,” and so on, but I do not take extreme positions on what is meant by those predicates (and have devoted much of the work in this dissertation to making clear the fairly uncontentious and deflationary ways in which qualia have these features). I shall briefly compare Dennett’s readings of the four properties with my own.

For qualia to be “private,” says Dennett, means that necessarily no interpersonal comparisons—such as functional or physiological ones—are possible (46). This is, apparently, because Dennett thinks qualia are necessarily not informational, dispositional, functional or physiological properties. I describe in some detail in what ways qualia can be said to be “private” in Chapter Six, and conclude that qualia are indeed, as a matter of contingent fact, privately possessed, privately apprehended with privileged but not necessarily authoritative access, and in some sense “perspectival” (in virtue of being phenomenal). But this certainly does not mean I adopt the view that interpersonal qualia comparisons are necessarily impossible; by contrast I even go so far as to suggest the possibility of qualia transplants, from one individual to another, or of two individuals somehow sharing the very same phenomenal brain state and so phenomenally apprehending the same qualia. And I explicitly countenance the possibility that, if supported by an adequately confirmed theory, third-person observations of a subject’s brain could sometimes even over-ride the subject’s sincere reports of first-person qualia apprehension. Finally, as explained in the previous chapter for example, I do not hold it to be impossible

---

5 I do not even hold that qualia are properties of conscious experience; rather, I have suggested, conscious experience is made up of qualia-tokens, which in turn are properties of the brain.
that qualia are identical with third-person observable physical properties.

Dennett writes that to be “intrinsic” (for qualia) is *inter alia* to be “somehow atomic and un-analysable. Since they are ‘simple’ or ‘homogeneous’ there is nothing to get hold of when trying to describe such a property to one unacquainted with the particular instance in question” (46). I have implicitly taken qualia to be intrinsic properties, but just adopted the minimal, and fairly standard, definition that some property is intrinsic iff it is not extrinsic, where a property is extrinsic “just if to possess \( F \) is to stand in some relation to other, wholly distinct or non-overlapping, contingent things” (Garrett 1995). I consider *atomicity* separately (Chapter Ten, section 1a), and far from holding that it is a necessary property of qualia, I claim that it is ultimately an empirical issue. Indeed, I suggest above that we had probably better assume for now that qualia are *not* atomic, since if they are ontologically atomic properties then third-person accounts of qualia will be much more difficult and perhaps impossible. Considerations something like this are presumably behind Dennett’s suggestion, above, that atomicity implies inef-fability.

I cannot see how atomicity is supposed to be implied by intrinsicality (since, after all, a raft of other putatively intrinsic properties, such as being cubic or liquid, are not atomic) except perhaps in combination with a Dennettian variety of privacy, such that qualia are both intrinsic (i.e. non-relational) and necessarily non-identical with any complex of physiological or functional properties. Under such circumstances, we might be able to say that qualia are of their essence unanalysable; however, of course, I hold no such view.

Qualia are “immediate,” for Dennett, again because they are *not* brain properties but are “essentially directly accessible to the consciousness of their experiencer (whatever that means)” (46).

Among the non-intentional (and hence qualitative?) properties of my visual states are their physiological properties. Might these very properties be … qualia …? It is supposed to be obvious, I take it, that these sorts of features are ruled out because they are not ‘accessible to introspection.’ (47–48)

---

6 Dennett too, in his discussions 68 ff., works with just this sort of contrast.
In Chapter Four above I distinguished between phenomenal and propositional apprehension and argued that qualia are “immediate,” on any plausible interpretation of the word, only to phenomenal apprehension. Qualia are *phenomenal* properties; the tokening of some quale is the sensation of that property: to be in a brain state with the quale *taste-of-milk*, for example, just is to undergo the phenomenal gustatory sensation of drinking milk, and in this respect we immediately apprehend—in one sense, “introspect”—qualia. On the other hand, as I have already pointed out, I am quite ready to suppose that qualia might be identical with certain “physiological properties” which are observable in some manner other than phenomenally, and even if they are not I still wish to insist that they are brain properties.

Finally, Dennett stipulates that qualia are “ineffable” in the sense that “one cannot say to another, no matter how eloquent one is and no matter how co-operative and imaginative one’s audience is, exactly what way one is currently seeing, tasting, smelling, and so forth” (46). More generally, we might say, one cannot know what it is like to have a certain sort of experience except by having such an experience oneself. For once, I agree that this is the case, in just the way that Dennett means it … but so, in fact, does Dennett! “Why does it seem that our conscious experiences have ineffable properties? Because they do have *practically* ineffable properties” (69). However, our reasons for asserting this ineffability thesis differ somewhat.

Dennett, in brief, argues that we possess the capacity to discriminate very complicated properties in the external world (such as the cry of an osprey, as opposed to that of a fish eagle), and that we often cannot project into the future exactly “what property my newfound property detector detects” (70). We therefore have two practical explanations for ineffability. First, some perceptions are “deliverance[s] of an informationally very sensitive portion of my nervous system” (70) and as such are extremely hard to describe; and second, since we do not know ex-

---

7 I do not follow Dennett, then, when he says that the “properties of conscious experience” are such “that one’s epistemic relation to them is *exactly* the same as one’s epistemic relation to such external, but readily—if fallibly—detectable, properties as room temperature or weight” (60).

8 This definition is also accepted and promulgated by Lewis (1980, 1988), Nemirov (1980, 1990), Seager (1991), Paul Churchland (1985) and Patricia Churchland (1986).
actly what property is being detected, we cannot furnish even a very complicated predicate to describe that property—“[t]he only readily available way of saying what property $M$ is is just to point to our $M$ detector and say that $M$ is the … property detected by this thing here” (71).

By contrast, I think that qualia are ineffable because they are not quasi-linguistic entities, nor replaceable by quasi-linguistic entities, but are properties of the brain. Qualia are incommunicable in just the same way that the property firing at 90 hertz or having sodium ionisation level $n$ are incommunicable. My describing what it is to fire at 90 hertz, no matter how silver my tongue or how enthusiastic my listeners, will not be a firing at 90 hertz. All that I can communicate verbally is information about 90 hertz firings, descriptions of such firings; similarly, though I cannot (verbally) communicate my qualia to you, I can describe them for you, in great detail if you like. Properties can, however, be communicated in the following way—they can be brought about in the recipient, in the way a battery ‘communicates’ electrical charge to portions of an appliance. Likewise, if I knew how to stimulate your brain, Penfield-style, such that you experienced the taste of strawberries, then I could ‘communicate’ that quale to you.¹⁰

My conclusion, to this point, is that the ‘theory-relativity’ strand of Dennett’s argument misses its target. The theory of qualia he presents for demolition is far more full-blooded than that needed to simply give the term an intension, and is arguably a far from canonical model even among more theoretically sophisticated variants. My own model of qualia bears a superficial resemblance to that Dennett attacks, but is sufficiently different to escape the brunt of Dennett’s complaints.

**b) The ‘Deep Structure’ Of ‘Quining Qualia’**

Although Dennett disingenuously insists that he is not giving “rigorous arguments” in

---

¹ It was once said that qualia are ineffable because they are “that which remains untouched and unaltered, however it is construed by thought” (C.I. Lewis 1929, 53). That is, only propositions or concepts are ‘effable,’ not qualia themselves.

¹⁰ There is some prospect that this kind of communication can in fact be performed aurally: certain strains of music can make us feel melancholy, some poems can produce clear visual images in the listener's head, and so on.
“Quining Qualia”—but instead merely proffering “intuition pumps” to help us shed our old pathogenic modes of thought (44)—and although he apparently targets what he calls the four characteristic properties of qualia for his main attack, in fact I think that Dennett offers us first and foremost a sequence of crypto-verificationist assaults on paradigm cases of qualia. This argument has the following form:

i) Paradigm cases of qualia shift (traditionally designed to isolate qualia as those properties which have changed, while all else stays constant) are not clear cut: there is always more than one possible interpretation of the empirical facts of the case, such that qualia may or may not have changed (or changed in isolation).

ii) Therefore, since even these paradigm cases fail to unambiguously pick out qualia, propositions about qualia cannot be justifiably asserted or empirically verified.

iii) Therefore, “qualia” has no denotation.

Dennett, then, is not merely arguing that “qualia” is a term which finds its meaning within a bankrupt theory, but more fundamentally that when we try and pin down what we are talking about, even pre-theoretically, it turns out that there is nothing there. This can be considered a second wave of assault, designed to back up the argument from theory-relativity: even if it is true that “ quale” does not mean anything so full-blooded as “whatever it is which is ineffable, intrinsic, private and immediately apprehensible,” but instead just means, as I defined the term above, “phenomenal colours, sounds, tastes, etc.,” it might still be that the term has no reference. Thus, Dennett asserts that there is no such property as “the way the juice tastes to Dennett at time t,” for example (45).

The first paradigm case for qualia that Dennett considers is the deliberately neurologically

---

11 Don Ross (1993) has made a similar analysis of Dennett’s paper; he notes that when explicitly laid out like this the argument loses some of its initial plausibility, which is possibly why Dennett is rather less than explicit.

12 “Otto can’t say anything more about the property he calls pink by saying ‘It’s this!’ (taking himself to be pointing ‘inside’ at a private, phenomenological property of his experience). All that move accomplishes (at best) is to point to his own idiosyncratic color-discrimination state …. Otto points to his discrimination-device, perhaps, but not to any quale that is exuded by it, or worn by it, or rendered by it, when it does its work. There are no such things.” (Dennett 1991, 383)
caused intra-personal inversion of qualia (50). Here, according to the qualophile, not only does it seem clear that qualia are \textit{those sensations which have changed}, but it even appears in principle possible to develop a well-confirmed neurophysiological account of qualia shifts. However Dennett objects that he can describe two possible ways in which the “qualia inversion” could have been brought about (50–51):

a) a reversal of the “early” qualia-producing channels (e.g. the optic nerve);

b) a reversal of the memory-access links that allow us to compare today’s qualia with those of yesterday.

In the second circumstance, the subject’s qualia have not in fact been changed at all. And, Dennett urges, the subject has in principle \textit{no way of telling} merely by introspection which of the two circumstances has occurred.

Dennett’s well-known second paradigm case is the example of Chase and Sanborn, the Maxwell House coffee-tasters (52). Chase complains that, over the years, he has grown to dislike the taste of Maxwell House coffee; Sanborn responds that, although he agrees the taste has remained constant over the years, and he too no longer likes to drink Maxwell House coffee, this is because his coffee-tasting system has broken down, and the coffee no longer tastes the same \textit{to him} (52–53). Dennett picks out three positions along a spectrum of possibilities for Chase and Sanborn:

a) their qualia have stayed constant, but their reactive attitudes (their tastes, etc.) have changed;

b) their qualia have altered gradually and imperceptibly over the years, but their standards of taste have not budged;

c) some predicament intermediate between these two poles.

Once again, Chase and Sanborn themselves are unable to distinguish between the different cases. Empirical tests (such as checking Chase’s abilities in blind testings, and his knowledge of the canons of coffee style), though they might be helpful in determining the combined \textit{outcome} of changes in qualia and judgements, are much less useful in resolving the issue of how much of each change has taken place.
Finally we have paradigm case three: Dennett imagines a case of post-surgical adaptation, where a patient who has undergone some form of qualia inversion gradually compensates so that after a while she says that the sweet thing once again tastes sweet, the sour sour, and so on (57). There will always be two ways of interpreting this change, Dennett says, no matter what the evidence. But here Dennett gives a new argument: even if we know precisely what physiological changes have taken place in the subject, we do not know where “to draw the line at which the putative qualia appear as properties of that phase of the process” (57). For any point in the neurophysiological processing, says Dennett, we have (at least) two possibilities:

a) we could refer to it as an input to the qualia phase; or
b) we could treat it as an output of the qualia phase.

We have no way of choosing between the two, Dennett says. And, ex hypothesi, the subject will be of no help in deciding.

How shall we respond to this extended crypto-verificationist argument? First, note that Dennett’s claim about the ambiguity in cases one and two is importantly restricted to the first person. It is only the subject of the qualia inversion, in these examples, who is unable to make justified assertions about the qualia changes. Dennett gives us no reason to believe that all third-person observers would necessarily be ignorant about which of the two or three circumstances had occurred: after all, the first inversion was ex hypothesi deliberately brought about by surgeons. Dennett’s third example does make claims about the third-person observer, but surely begs the question: why should Dennett think that he has the license to pronounce that no future neurophysiology will be able, even roughly, to identify the material states that instance qualia, or to distinguish between the brain areas of perceptual awareness and of memory access and comparison?13

13 “Suppose that the evidence for the sensory-vector story turns out to be overwhelming. Now suppose we have charted the history of Chase’s taste buds and have discovered that over time there have been small differences at the taste receptors on the tongue (destruction caused by coffee), causing a shift in the vectors as viewed on PET scanners. The vectors once looked like those of others who reported that ‘Maxwell House tastes great.’ Now they look like those of people who say they don’t like the taste of Maxwell House. … The inference to the best explanation is that his taste buds have undergone small changes that cause Maxwell House coffee to taste bad, whereas
So Dennett does not establish with these examples that claims about qualia are scientifically or objectively indeterminate and empirically unverifiable; at most he has argued that claims about qualia are introspectively unverifiable. But presumably, any verificationist-style move would have to depend upon establishing the former claim.¹⁴

Secondly, even the weaker introspective claim depends upon good old fashioned sceptical considerations to carry it through, and these considerations apply too generally: they are reasons, not just for us to be dubious about qualia in particular, but for various varieties of radical scepticism. The first consideration is the claim that, if we have experience only of the effects it used to taste good. … His memory is mistaken. Maxwell House coffee tastes different now than it used to.” (Flanagan 1992, 76)

¹⁴ Something similar is true of Dennett's discussion of the colour phi phenomenon (in 1991, 1992), though here it is qualia (a.k.a. 'conscious' awareness) and perceptual discriminations (as opposed to judgements of taste, as in the “Quining Qualia” examples) which are being blurred together. The colour phi phenomenon occurs when, for example, a green light (at point B) flashes very shortly after a flash of red light at point A; subjects report experiencing a coloured dot moving from A to B and changing colour from red to green halfway along its trajectory. At this halfway point, then, subjects claim they see a green dot before that green dot has actually appeared—indeed, before they can know that the green flash will appear. Dennett claims that there is in principle no difference between the following two scenarios: subjects experience a lingering red quale until the green dot appears, but then subjects misremember their own qualia and report having experienced the dot changing colour earlier (the “Orwellian” story); or their experience is delayed (by up to 200 milliseconds or so) after their perceptual discriminations, and the whole quale sequence is constructed after the green dot has appeared. Dennett writes that both models can account equally well “for all data—not just the data we already have, but the data we can imagine getting in the future” (Dennett and Kinsbourne 1992, 158). But this would be so only if we already believed it impossible to get any data about qualia beyond first-person verbal reports. As Ned Block commented at the time, “Dennett and Kinsbourne’s … argument hinges on the unmentioned and unargued assumption that there is no such thing as phenomenal consciousness” (1992, 175).

Dennett does consider the possibility that brain states can “become conscious” “by acquiring some property or by having the intensity of one of its properties boosted above some criterial level” (1992, 167), but objects that “[t]he actual, objectively measured simultaneities and sequences in this broad field are of no functional relevance unless they can also be accurately detected by mechanisms in the brain. What would make this sequence the stream of consciousness if the brain could not discern the sequence?” On my account, however, the properties in question—qualia—are explicitly phenomenal, and that’s what makes them the stream of consciousness; there is no need for them to be ‘detected’ elsewhere in the brain, and certainly not for them all to occur at one particular area of the brain, a 'consciousness module.'
of a causal sequence, then we cannot know which of various possible causes was the actual cause—thus our beliefs about our own qualia can be changed either by changes in the qualia themselves, or by changes in some other part of the causal chain bringing about those beliefs. The second is Kripkensteinian scepticism about memory—15—if our memory is fallible then any judgement or activity which requires comparison of the present with the past will also be fallible. The problem is that these sceptical considerations do not merely affect judgements about *qualia* but also, classically, things like the perception of the external world and the possibility of following a rule. For example, if we have direct experience of our own perceptions, then we cannot tell whether they originate in the world or Descartes’ *malin génie*. Similarly, for instance, if we cannot reliably recall how we have applied the rule “plus” in the past (Kripke 1982), we cannot be sure that we are applying it consistently or correctly now. Since qualia are no more vulnerable to these kinds of attack than perception and rule-following, they should be called into no more doubt than is cast upon perception and rule-following.

Thirdly and finally, that two things are empirically difficult to distinguish does not show that one exists and the other does not. That our conceptual resources might be hazy on boundary cases does not show either that the concepts have absolutely no content nor that the referents of these concepts do not exist. That various possibilities are empirically hard to disentangle does not show that getting straight about what these possibilities actually involve will be fruitless. Thus, even if Dennett were persuasive about these three cases, it is not clear that his eliminativist conclusion would follow through.

### 2. DRETSKE’S EXTERNALISM

In his recent book *Naturalizing the Mind* (1995), Fred Dretske applies his sophisticated version of functionalism (which he calls naturalist representationalism16) to phenomenal experience.

---


16 Dretske has developed and refined his version of functionalism over many years: see in particular his 1981 and 1988.
His general thesis is that all mental facts are representational facts, and all representational facts are facts about informational functions (1995, xiii): that is, in a nutshell, that all mental facts are facts about what information some representation is designed to convey. In chapter three of *Naturalizing the Mind* Dretske applies this general rule to qualia with the result (in his case at least) that he develops an externalist account of qualia.

*a) The Roots of Dretske’s Externalism*

Dretske identifies qualia with the properties that the external, objective objects of perception are, in a particular systemic way, represented to us as having (65). It follows that for Dretske qualia are properties of external objects of perception. For reasons that will become clear, however, Dretske pays scant attention to the nature of these properties; he focuses instead upon giving an account of the particular way in which qualia—these external properties—are represented to us. Not every representation picks out a quale; only certain types of mental representation do so, and it is these special kinds of representation that Dretske strives to characterise.

On Dretske’s account, a system, $S$, represents a property, $F$, iff $S$ has the function of providing information about the $F$ of a certain domain of objects. $S$ does so by occupying different discrete states, $s_1$, $s_2$, … $s_n$, corresponding to the different determinate values, $f_1$, $f_2$, … $f_n$, of $F$ (2). Dretske’s favourite example of a straightforward representational system is a speedometer: here $F$, the property represented, is speed; the domain contains only the vehicle to which the speedometer is attached; and the relevant discrete states of the system are positions of the needle on the dial.

Dretske’s notion of a function in this context has a strong teleological bent: what property a system represents depends not upon what properties it carries information about, but upon what property it is *intended* to represent (4). For example, a speedometer carries information (after an appropriate mathematical transformation) about axle rotation and wheel revolutions per minute; but it only represents the property of speed. This teleological component, Dretske
explains, allows for the possibility of misrepresentation: an object can retain a function even when it fails to perform it.

It is important to Dretske’s project that we accept that there are natural, as well as conventional, representations, whose ‘intended’ functions are derived from their evolutionary history or through individual learning (7): all mental states, Dretske says, are natural representations. Not all natural representations, however, are mental: animal homeostatic and immune systems, for example, plausibly utilise natural representations but would not normally be said to have mental states. Dretske does not, as far as I can tell, provide a full account of this difference in Naturalizing the Mind, though he does note that natural representations which do not directly supply information to a cognitive behaviour-control system are not mental (8 n.6, 19–20).

Dretske’s next move is to cash out the distinction, within the class of mental representations, between sensory and conceptual representations: between sensations of k’s blueness, for example, and judgements that k (or something) is (or looks) blue. For Dretske this distinction corresponds to another: that between systemic and acquired indicator functions. In the former case some state of the system regularly corresponds to some property that it is intended to indicate, and so systemically indicates it. In the latter case, some state of the representing system is assigned a particular representational content, independently of what it may actually systemically indicate. Dretske’s example is of a simple speedometer mechanism that represents vehicle speed by measuring the rotation of the axle. Some state $\beta$ of the speedometer dial systemically represents an axle rotation rate of $N$ rpm. In cars with different tire diameters the dial would be calibrated differently so that in one car state $\beta$ might have the acquired function of indicating 50 kph and in another 60 kph (12–14).

Dretske, then, claims that one instance of the difference between systemic and acquired representation is that between sensations and thoughts—between seeing a blue thing, and seeing it

17 Another of Dretske’s examples involves a pressure gauge whose needle varies systemically with air pressure; such instruments are routinely calibrated to show altitude in, say, metres above sea level, and a given pointer position will then, Dretske says, be both a systemic representation of pressure and an acquired representation of altitude (20).
as blue. Though Dretske does not enlarge on this example, his text suggests the following scenario. Our visual systems have evolved to *systemically* indicate the determinate value of some objective determinable, which we call the property of colour—light-reflectance within certain illumination conditions, perhaps. This system has in fact been ‘designed’ to distinguish between more than 16 million determinate states (that is, different shades of colour). Our visual systems are calibrated to a greater or lesser degree of specificity among different individuals, and have the *acquired* function of indicating, presumably, the named hues, such as maroon, aquamarine and buttercup yellow.18

Dretske still has one more distinction to make. Not all natural systemic representations are *experiences*,19 on his account; experiences make up the proper subset20 of such representations that “service the construction of” acquired representations that can be calibrated to “more effectively service an organism’s needs and desires”—that is, they “are the states whose functions it is to supply information to a cognitive system for calibration and use in the control and regulation of behavior” (19). Dretske once again provides a speedometer analogy: in a speedometer, those states “available for use in the control of behaviour” are the indicator states of the speedometer’s dial. In the simple device already described, the system’s analogue of experience is its representation of axle rotation speed; its belief-analogues are its representations of

---

18 It is tempting to object at this point that, since on Dretske’s theory only natural, systemic functions pick out qualia, it follows that “we are incapable of experiencing—of being phenomenally aware of—a variety of modern-day artefacts and properties. Since, e.g., there were no automobiles around when our perceptual systems were being ‘designed’ by natural selection, no ancestral perceptual system could have been selected for providing information about automobiles, and hence no natural representation could (presently) have the systemic function of indicating autos” (John Baker, personal communication). I think that this objection, when put this way, is mistaken: Dretske has in mind our natural, perceptual ability to discriminate between various different car-sized objects, and in this I think he is correct. A Cro-Magnon man could presumably perceptually discriminate a Toyota Tercel from a Chevrolet Cavalier even though, obviously, he would not have the slightest (conceptual) idea what they were. However, I do think worries of this kind conceal a deeper problem, that I attempt to bring out in my objection c) below (see also the discussion of the Myth of the Given above, Chapter Six, section six).

19 This is Dretske’s term, and he may well be using it in a way which is not compatible with my technical usage. Nothing much hangs on this for this discussion, as far as I can see.

20 It is not clear whether Dretske believes all, or only some, members of this set are experiences.
vehicle speed. In a more complex speedometer device, information about axle rotation is combined with information about the height of the axle above the road surface in order to determine the state of the dial. Here, the system’s ‘experiences’ and ‘beliefs’ are both about the speed of the vehicle, and information about axle rotation is ‘lost’—it is analogous to non-conscious information carried during the earlier stages of the processing of perception (such as the “primal sketch,” the primary visual area’s distorted map of light intensities at different parts of the retina).

Sensations or experiences, then, are in Dretske’s account systemic, natural representations that underlie the construction of behaviour-regulating acquired representations. Although Dretske is never fully explicit on this point, it seems by far the most plausible in the light of the text to assume that he sees qualia as the properties so represented, and so as distinct from sensations and experiences (and their properties). As he puts it, “the Representational Thesis identifies the qualities of experience—qualia—with the properties objects are represented, as having” (65).

b) What Is it Like to See a Poodle?

So, given this picture, what can Dretske say about the phenomenal qualities of conscious experience? For example, for Dretske, what is it like to see a bat … or, to use his own example, a poodle? Though it is not clear that Dretske ever really answers this question—which is surely the one of most interest to the qualophile—he does address two related questions:

1) When does something look like a poodle to someone? This question he seems to interpret as meaning: What is it for some organism to have a specifically poodle-like experience (rather than an experience of some other sort)?

2) How do things seem to someone experiencing a poodle representation?

His answer to the first question is that something looks (sensorially, rather than conceptually) like a poodle to $S$ if it looks the way poodles normally look to $S$ and if it looks different to $S$ than some proper contrast class (e.g. other dogs). As Dretske puts it, $S$’s “experience of the dog represents the dog … as having … the manifest properties of poodles, those proper-
ties that make poodles look so much different from other dogs” (67). An organism is capable of having poodle experiences only if it has a visual system that has been designed to discriminate between medium-sized objects at the appropriate level of detail.

Dretske recognises that possibilities similar to that of inverted qualia still apply to this account: that $S_1$ and $S_2$ are equally good (or bad) at discriminating poodles does not imply that poodles look the same to them. Thus, for example, two people might be equally unable to distinguish between red wines, but one might experience “an exquisite Burgundy taste” when they drink any red wine, and the other might always experience “the taste of cheap Chianti” (74–75). Dretske does not consider this a problem for his account, he says, because he agrees that “the qualitative character of perceptual experience … is not functionally definable” (72). For him, as we have seen, qualia are the properties that an experience has the function of systemically representing something as having, whether it is actually performing that function or not: thus (i) these properties need not “express themselves” in the behavioural dispositions of the system “in which they exist” (72); but (ii) these properties are objectively determinable by the following set of identities:

a) qualia = experienced properties;
b) experienced properties = systemically represented properties;
c) systemically represented properties = those properties about which the senses have the natural function of providing information (72).

Thus, according to Dretske, questions about qualia are really questions about the properties systemic representational states of a system have the function of indicating. (Whether this is a satisfactory reply to the inverted qualia problem I shall consider in a moment.)

What about question two: How do things seem to perceivers? Dretske suggests that the way an experience represents an object is the way that object would be if the representational system were working correctly (73). Thus, if some organism is designed to differentiate between poodles and other objects, and that organism hallucinates that all the medium-sized objects around her are poodles, then everything seems to be a poodle to her. As we have seen, the design of the system—what it would be for the system to “work properly”—is essential for
Dretske. Thus, even if two speedometers have identical functionality, if one was originally more discriminating and reads 78.00 kph (at speeds between roughly 77.50 and 78.50 kph) then it has a different quale—a.k.a., for Dretske, means something different—than the other which reads merely 78 kph over the same range of speeds (pp. 76–77). The latter does not, and could not ever, experience 78.00 kph.

It follows from this that:

Knowing what bats, fish, and neighbors experience is, in principle, no different from knowing how things ‘seem’ to a measuring instrument. In both cases it is a question of determining how a system is representing the world (81–82).

For example, Dretske imagines a mono-representational parasite that only has a thermal sense, with which it picks out receptive hosts that have a body temperature of exactly 18°C. To know what it is like for that parasite to sense a receptive host,

[all you have to know is what temperature is. If you know enough to know what it is to be at a temperature of 18°C, you know all there is to know about the quality of this parasite’s experience. … For, if things are working right, what the host is—18°C—is how things seem to the parasite. So if you want to know how things seem to the parasite, look at the host (83).

In addition to deriving this position from his general framework, Dretske gives what he takes to be an independent argument for this conclusion:

i) Qualia are supposed to be the way things seem in the sense modality in question.
ii) Things sometimes are the way they seem.
iii) Therefore qualia are exactly the properties the object being perceived has when the perception is veridical.
iv) The quale of the parasite is just like that it has when its perception is veridical.
v) Therefore the quale of the parasite has to be exactly the property the object has, i.e. 18°C. (83–84)

There is no more to experiencing blue than there is to the (objective, external) colour blue since the colour blue is the colour one experiences, Dretske writes; “there is no more to experiencing an electric field of type T [if you are, say, a dogfish] than there is to being an electric
field of type \( T \)" (85). In fact, it turns out, there is considerably less: all that a dogfish represents about the field, says Dretske, is its geometrical configuration, not that it is an electrical field, that it is a force generated by electromagnetic charge, and so on. If one knew everything about what properties bats and dogfish systemically represent to themselves, then one would know what it is like to be a dogfish or a bat (94). Further, Dretske insists, one can know what it is like to have a certain experience without being able to discriminate that property yourself. The dogfish is one example; Mary the colour-blind scientist in Frank Jackson’s well known thought experiment (1986) another. Dretske even claims that one could know what it is like to hear a change of musical key without being able to discriminate one yourself, as long as one knows what it is to change key (85–86).

Here, then, in summary is how I understand Dretske’s account of qualia. Qualia are completely characterised as the objective, external properties that those mental representations which are our sensations have the natural, systemic function of indicating. Therefore sensations are of their associated qualia—those objective properties it is their systemic function to pick out—and are just as objectively specifiable as the systemic functions of any physical system.

c) Five Objections

Dretske suggests, at the beginning of his book, that his Representational Naturalism “is the only approach to consciousness that has much to say about the baffling problems of phenomenal experience” (xiii). And indeed, Dretske’s account has quite a lot of initial plausibility. Unlike proponents of more traditional functional analyses he seems ready and willing to deal with the “inverted spectra/earth” or “knowledge argument” families of counter-examples, and he makes moves intended to prevent his theory being too functionally “liberal.” Furthermore, unlike many of the opponents of this kind of naturalistic analysis (and unlike me), Dretske would face no so-called “explanatory gap”—no serious problem in explaining how qualia can

---

21 I hope my own theory is a counter-example to this claim.
22 This term was popularised by Ned Block in his “Troubles with Functionalism” (1980).
be related to, and explained in terms of, scientific theories of the physical world. Nevertheless, I think Dretske is wrong.

i) Dretske’s non-standard usage of qualia.

Dretske’s characterisation of qualia is radically non-standard. Far more than does my account, his externalist usage of the term flouts standard conventions or assumptions about qualia. First, standardly, qualia are mental properties. One usually speaks of qualia as being part of the mental life of the experiencer: organisms either “have” or “do not have” qualia, we say. Dretske, though he does not mention this, is committed to the position that qualia are non-mental, since he claims both that only representations are mental and that qualia are not representations but what is represented—external properties in the world. It follows that qualia are not in any way properties of their experiencer—it no longer makes any sense to say a conscious being “has” qualia, on Dretske’s account: external objects are what “have” qualia.

This also conflicts with the usual assumption that qualia are putatively (that is, if they exist at all) properties of conscious mental states. As we have seen, I also challenge elements of this pre-supposition, but from a much less extreme standpoint than Dretske. I accept that qualia are, in some sense, properties of our mental representations—I just deny that these representations are anything other than the relevant states of the brain as these states are already more or less understood (rather than being additional mental objects). Dretske, by contrast, treats qualia, not as properties of our representations, but of the objects of those representations.

Dretske’s non-standard (and non-explicit) usage of the term “qualia” is problematic because it tends to foster confusion in the reader; this is especially true since in many ways, Dretske’s notion of sense experience—roughly, the indicative states typically brought about in us by external properties (see the next section)—is much closer to the standard definition of ‘qualia’ than is his account of qualia, which invites a puzzling conflation of the two. In fact what he calls “qualia” are a proper subset of what I call simply properties of external objects, specifically

---

23 Qualia can be properties possessed by both mental states and experiencers, at least loosely speaking, just as digestion is a property both of digestive systems and whole organisms.
those external properties which have the power to create in us sense experiences; and “sense experiences,” for Dretske, seem very similar to what I have been taking qualifiers to be—that is, the states of which qualia are properties—this reversal, on Dretske’s part, can seem very misleading. Also, because of the approach Dretske takes, his comments sometimes seem only tangential to the traditional qualia problems, such as the possibility of absent qualia and the apparently non-physical “subjectivity” of qualia. However, that Dretske’s account is non-standard does not by itself mean it is wrong. The following objections amplify upon the reasons why I think Dretske’s externalist approach fails.

ii) The problem of demarcating the mental.

For Dretske, virtually any property (being an alligator, being a haircut, being larger than a bread-box) can in principle be a quale. The real work in Dretske’s account is being done by the notion of a “mental experience”: in his view qualia are merely the objects of such experiences. However, Dretske’s account of mental experiences seems too loose and thin to bear the added philosophical weight the concept must now carry. For Dretske, anything, s, that satisfies all the following conditions is a sense experience:

a) s is a discrete state of a representational system with the function of indicating some determinate value of an objective determinable;

b) s’s indicator function is a natural one;

c) s’s indicator function is a systemic one;

Indeed, it is possible, and perhaps more comfortable, to interpret Dretske as not making the identity claim about qualia that I claim he is making, but instead saying something like the following: to have qualia of type T is nothing more than to token states that systemically represent something as having property T; seeing red is merely to have a sense experience of (i.e. whose function is to pick out) red, and nothing more. But where then, on this account, are the qualia—the phenomenal properties of experience—exactly? What are they properties of on this interpretation? There seem to be three options. Either Dretske is an eliminativist about qualia (which he denies, and which is anyway rather uninteresting); or qualia are additional, as yet unmentioned, properties of sense-experiences (in which case they have yet to be described, let alone theorised about); or they are properties of the objects of sense-experience (which is how Dretske explicitly describes them, and which takes us back to my interpretation).
Chapter 11: “Three Central Counter-Theories.”

262

d) is “mental” in some incompletely specified sense that includes underlying a system of acquired representations which are used to control behaviour.

These liberal constraints are in danger of committing Dretske to the position that entities unanimously considered non-sentient experience qualia. Consider a simple plant that continually sucks up water through its root system until the pressure threatens to burst its cells, and then opens pores on its leaves which allow the fluid to transpire; when the pressure in its cells has fallen to a satisfactory level, the pores are closed again. Such a system apparently satisfies Dretske’s conditions: on his account, the pressure inside the plant’s cells constitutes a quale that brings about systemic changes in the system (changes to the plant’s cell walls), which is ‘calibrated’ to either believe that the pressure is too high or that it is not, and adjusts its behaviour appropriately. That is, the state of the cell walls is a natural, systemic indicator of the pressure inside those cells—each discrete state of the walls represents a different pressure level (criteria a)–c)). For reasons of natural teleology the plant has been ‘calibrated’ to respond to certain categories of these systemic indicator states, and adjust its behaviour appropriately: that is, it ‘experiences’ the acquired representations too much pressure or acceptable pressure (criterion d)).

Dretske must presumably reject this counterexample by claiming that, once condition d), mentality, is cashed out in more detail, plants will fail to satisfy this condition. That is, much more is needed in a satisfactory account of Dretske’s condition d) than the assertion that “supplying information for calibration and use in the control and regulation of behaviour” is the hallmark of the mental. At best, this unclarity at the heart of Dretske’s account is unfortunate—Dretske is, after all, engaged in the project of providing a theory of the mind. At worst, this is circular: only entities with a (conscious) mental life are candidates for possessing experience, where what it is to have a (conscious) mental life is classically understood as being an experiencing subject.

Perhaps it is best, therefore, to treat Dretske not as providing a theoretical treatment of our extant notion of qualia—despite textual indications that this is the case25—but as giving a prin-

---

25 For example, Dretske expresses the intention of trying to solve “the baffling problems of conscious experi-
cipated *redefinition* of the term such that plants, it turns out, *do* have qualia and that is not a problem. However, if this is so, then it is far from clear how defensible Dretske’s conditions might be *in themselves* (rather than just as an attempt to provide a definition roughly coextensive with what we mean by qualia). Apart from the vagueness of his “mentality” condition, it is difficult to see how condition b), that the indicator function be a natural one, can now be motivated.

Why should it make a difference exactly what roots the teleology of a function, as long as it has one? Suppose some advanced race were able to build or replicate living organisms by manipulating molecular raw materials, and imagine that, by chance, their designers hit upon a form identical in every relevant physical way with a human baby. Why should we say that this baby, when grown to adult-hood, does not experience perceptual sensations in the same way that we do? After all, *ex hypothesi*, it interacts with the world in exactly the same way we do, and precisely similar events take place in its brain.26

But if we remove condition b), while retaining the minimalist version of condition d), then Dretske’s class of beings able to experience qualia inflates drastically to include, for example, mechanisms like cars and thermostats.27 Consider a system where discrete states of a speedometer’s dial systemically correspond to the speed of axle rotation. Suppose that the states of the dial are ‘read’ by an on-board computer, which has been calibrated to correlate these states

---

26 A similar thought-experiment can be levelled against Dretske’s claim that any teleology is involved at all. Suppose, *per impossibile*, that our human child were not designed, nor evolved, but created instantly from molecular raw materials by a freak lightening storm on some distant planet barren of sentient life (though not inhospitable). Again, the unfortunate baby would have the same causal connections to the world that we do, and would pass through brain states of the same type as ours—increased activity in striate cortex areas V1 and V2 after input from the Lateral Geniculate Nucleus (LGN), for example. Yet Dretske would have it that, because of the accident of its birth, this organism undergoes no sense experience whatsoever; that all is dark inside its head, while ours—physically identical—is alight with visual, aural, tactile phenomena. This seems deeply implausible, and a violation of the principle of the supervenience of qualia. However, since Dretske explicitly addresses this point in his book, and believes he has a reply to it (141 ff.), I shall not pursue this line here.

27 Despite Dretske’s surprisingly general claim that “g gauges and instruments do not do anything with the information they provide” (167).
with acquired representations of the speed of the vehicle; and imagine that this computer uses these acquired representations to influence its control over the car’s behaviour—its speed, gearing, traction, and so on. Now, do we really want to say that a late model Toyota has a mental life, or do we prefer to insist that having qualia is separable from having a mental life? Neither position is attractive.

iii) The problem of individuating qualia.

We have seen that Dretske identifies qualia with the objective properties whose determinate values our experiences have the function of discriminating. However, as Dretske admits, it is not always easy to identify those properties: it is not always possible to determine the function of our experiences. Dretske takes this to be at base an empirical problem (88 ff.): thus, identifying the objective property of colour is straightforwardly a matter of discovering what property in the world our relevant visual apparatus evolved to indicate, although this may since have become confused by the phenomenon of metameric matching\(^{28}\) and so on.

However, there are reasons to believe that this uncertainty is actually a conceptual problem with Dretske’s account. First, though there may be a finitely describable set of determinate physical conditions which bring about every instance of an experience of the colour red, these physical conditions, contra Dretske, still need not constitute an objective external physical property. As we saw above (Chapter Four, section two), it is not implausible that our sensations of redness—and not just our judgements of redness—are influenced by psychological factors, such as our expectations and other tacit beliefs, which affect pre-conscious visual processing. The relevance of this to Dretske’s account is that the evolution of our visual systems may have ‘taken this into account’: we might have evolved to detect human-coloured things, not just through detecting some objective physical property of human skin, but also through assessing whether the context makes them likely to be human, and abstracting away from the actual physical property to filter out conditions of illumination, for example. If this were so, then on

\(^{28}\) That is, the phenomenon where a wide variety of objective circumstances can give rise to the same colour experience.
Dretske’s account the property our pinkish-brown colour experiences would have the natural, systemic function of indicating would be something like “what might be expected to have human-skin colour,” rather than any objective external property surface reflectance properties.\textsuperscript{29} The cognitively laden character of such properties would not only require revision of Dretske’s account, but would play havoc with his central conclusions about the objectivity and third-person accessibility of qualia.

The second difficulty with individuating the objective properties that are supposed to be qualia is that, within Dretske’s account, there is no principled way of saying just which qualia some reasonably normal, functioning human beings are experiencing under certain conditions. Dretske says that someone has a visual experience of a poodle only if their visual system has the function of demarcating between poodles and everything else—that is, if the representation has the function of indicating poodle-hood. But what is the relevant contrast class here? What is the property of looking like a poodle? Dretske admits that very good fakes, such as woolly robot poodles, do have that property, but insists that blurry medium-sized blobs do not—if all you can see are blurry blobs, then nothing looks like a poodle to you (66 ff.). But what about bichon frisés? These are small woolly dogs that, one is tempted to say, look like poodles. Suppose you cannot perceptually distinguish between poodles and bichon frisés (because you have less than 20/20—just a shade blurry—vision, for example): does this mean that you in fact do not have poodle sensations, but instead small-woolly-dog sensations? Presumably, if any normal human’s visual system has the function of distinguishing poodles from bichon frisés, then yours does too: that, for Dretske, means you are capable of experiencing both poodle qualia and bichon frisé qualia. Which of the two qualia do you experience on this occasion, then … or is it some third quale altogether?\textsuperscript{30} Dretske’s model gives us no way of deciding.

Actually the confusion is even worse than this: many normal human beings can tell the dif-

\textsuperscript{29} Information about the putative objective external property of pinkish-brownness would be ‘lost’ before the point of mental representation, analogously with representations of axle-rotation speed in Dretske’s more complex speedometer system (see above).

\textsuperscript{30} Dretske mentions this very situation on page 69, but does not address this problem.
ference between miniature poodles and toy poodles, or between their poodle Fargo and the poodle next door, George. So why not a “my poodle Fargo” quale? Worse still, our visual system has the function of being able to spot differences in objects over time. So do we have the quale—that is, the objective property—“my poodle Fargo shortly after his annual bath”? In fact, if we are to take seriously Dretske’s model of experiences as being pre-conceptual representations of those determinate properties that our perceptual systems have the function of discriminating, then there are almost as many visual qualia as there are moments of visual perception.  

This massive explosion in the number of qualia, together with the prospect of their extreme transience and observer-relativity, casts dark shadows over Dretske’s cheerfully hard-headed picture of qualia as stable objects of objective scientific scrutiny.

iv) The problem of intentional inexistence.

Dretske holds that qualia are the external, objective properties that sense experiences have the function of representing. However, as Dretske notes, misrepresentation is possible: the world need not always be as it is experienced (4). Something can look blue but actually be some other colour altogether—this is the phenomenon upon which the argument from illusion is built, and the consequences for externalism have already been considered above (Chapter Three, section 2b). In brief, the difficulty is to say, in such cases, what entity is phenomenally blue.

On Dretske’s picture, the quale _phenomenal blueness_ is just the same objective property as ‘ordinary’ blueness (whatever exactly that is) … it is a physical property held in common, let’s suppose, by cornflowers, a clear sunlit sky and lapis lazuli. It is not a property of mental repre-

---

31 If we treat Dretske’s model as identifying qualia with determinables, then the problem reshapes itself as being one of bundling the determinate values together into the appropriate determinables, so to speak. So, for instance, is my experience of Fargo a “my dog” quale, a “small woolly dog” quale, a “poodle” quale, or just a “dog” quale? Again, there is no obvious principled way of deciding.

32 Not that my own account is necessarily much better in this respect; rather, I wish to show it is at any rate no worse off. My model, at least, is under no obligation to correlate discrete properties within the sensory manifold with particular functions.
sentations; it is, rather, indicated by such representations. Yet it is possible to have blue qualia
where there are no blue objects … where there is nothing that has the objective property of
blueness. Suppose, for example, that one is gazing fixedly at a large orange screen after just
looking at a bright blue light, and that a blue after-image is swimming across one’s gaze.

What can Dretske say here? It seems incoherent to assert that qualia are nothing more than
properties of external objects (like skies and flowers), and that this is a case of misperception
where there is no external object which has that property in sight, and that there is currently an
instance of that property present. The only consistent alternative for Dretske is to assert that,
during cases of misrepresentation, no instance of the represented property is in fact present.
The highly unusual conclusion follows that, while perception involves qualia, misperception
does not. And this conclusion also leads us into an unpleasant dilemma. Either Dretske must
finally bite the bullet and admit that qualia (in his sense) have nothing to do with what mental
life feels like, in which case he still owes us an explanation of the subjective, phenomenal quali-
ties of consciousness and has not dealt with the qualia problem at all; or he must assert that, by
contrast with veridical perception, there is nothing it feels like to misperceive (or at least that it
feels very different from veridical perception). In short, either the absence of qualia makes a
difference to subjective feel, or it does not, and either way Dretske’s theory is unpalatable.

v) The conflation of representations with what is represented.

It seems abundantly clear that there is no way of introducing an appropriate phenomenal ele-
ment into Dretske’s treatment of what he calls “qualia” themselves: qualia, for him, are just
regular, everyday properties that happen to be the object of certain sorts of discriminations.
But perhaps we would have more luck with Dretske’s account of “experiences”: after all these
are conscious mental states, even on Dretske’s account; they are representations of other prop-
erties, just as qualia and their ilk were classically supposed to be. The way here, however, is
blocked by Dretske’s absolute three-way identification of what it is like to have certain experi-
ences with the content of these representations with that quale they have the function of repre-
senting.
On almost any theory of experience—certainly on a representational theory—there is no more to the quality of one’s experiences in experiencing blue than there is to the color blue since the color blue is the color one experiences. Likewise, there is no more to experiencing an electric field of type $T$ than there is to being an electric field of type $T$ since $T$ is exactly what makes this electric field experience different from other experiences of the electric field. $T$ is the quale of this experience. (85)

The content and feel of an experience of an electrical field is, apparently for Dretske, identical with the property of being an electric field itself—or rather, being a little more careful about it, a description of the quality and content of an experience is exhausted by statements to the effect that an objective determinable is one way rather than the other. On this account, as Dretske himself points out with satisfaction, one can learn about what it feels like to be a dogfish experiencing lines of electric charge in the surrounding water by discovering more about electrical fields—how they work, how they are shaped, and so on (81 ff.). A certain dogfish experience, then, can presumably on Dretske’s theory, be exhaustively described by stating that the electrical field surrounding it has some shape $x$ and some charge $y$.

As a general claim about representations, even as teleologically understood, this simple identification is surely false. Even if one identifies the content of a representation with what it is intended to represent, one cannot identify this construct with the way the representation is configured, or looks or feels. Consider the case of a water-colour painting of a landscape in the Lake District. Suppose, for the sake of argument, that the content of this painting is precisely what it is intended to represent: say, that particular section of scenery plus a certain melancholy emotional mood. It is far from clear that, even if we were to know everything there is to know about the topography of the Lake District and about melancholy emotions, we would know what the painting looks like. This is even more strikingly evident if we suppose the painting to be in some kind of post-modern neo-cubist style, or if we change the example slightly to a descriptive passage with the same content but written in Armenian. The basic point here is simple enough. Various quite different states may represent some particular content $C$. And exhaustive knowledge about $C$ would not constitute—or even justify inference to—knowledge
about how C may be represented.33

As a claim about conscious, phenomenal representations in particular, there are strong reasons to think that Dretske’s conflation of representation with represented is just as erroneous. Two representations could have the self-same function of representing some object k as bearing all and only the members of some particular set F of properties, and yet those two representations could feel different. Knowing everything that science has to say about all the members of F does not by itself license an inference to what the sense-experience of k feels like for some other organism.

This sort of objection is commonly phrased in terms of the possibility of inverted qualia; because of Dretske’s idiosyncratic usage of the term “qualia,” for the purposes of this section I shall label this selfsame objection the possibility of “inverted raw feels,” where “raw feel” is merely a placeholder term for the way a mental state feels to its experiencer, however this is eventually cashed out. Dretske takes himself to have provided an account of raw feels that is complete in that it is not subject to the inverted raw feel objection: the objection that, for some theoretical identification of raw feels with some other set of properties P, the theory fails because raw feels can be varied while the members of P are held constant. But in fact Dretske’s theory is just as subject to this problem as are ‘mere’ functional or behavioural theories.

Suppose two systems have a taste-of-red-(as-opposed-to-white)-wine-detecting-mechanism; as Dretske points out, they could function equally well, but one could experience the taste of every red wine as being like what I experience when I sample a fine Burgundy, and the other could taste like what I experience when I sip a poor Chianti (71). Although Dretske is not fully explicit here, it seems clear that he intends to reconcile this possibility with the prohibition of inverted raw feels by insisting that only systems that have the function of detecting fine Burgundy can experience that quale/raw feel, and likewise with poor Chianti. Thus, one situation that could give rise to the situation described above would be if two well-calibrated red wine tasting machines both break down and their pointers get stuck at just one position in the red

33 And possibly vice versa: c.f. the problem of ‘inference’ to the external world.
wine ‘space,’ as it were, with one stuck at fine Burgundy and the unlucky one stuck at bad Chianti. On this account, Dretske will say, there is a representational difference between the two systems—it is not the case that raw feels have been altered without appropriate functional/representational changes—and so his identity thesis survives.

But Dretske has still not eliminated the possibility of undetected raw feel change. The only way for him to do this would be for him to insist that two raw feels are identical whenever they are the same representation of some objective property by two systems which are identically calibrated. Imagine two wine-tasting systems that are both broken down in exactly the same way: to them both, all red wine is represented by an experience that has the function of indicating fine Burgundy. If it is still conceptually possible that these experiences could feel different to the two systems, then the possibility of inverted raw feels remains. Dretske does hint that he might want to deny the possibility that two representations with identical functions can nevertheless differ (71, 75), but he never does so explicitly. And it is rather hard to see how he could: surely, wine-tasting machines with identical teleological functionality, designed to identify exactly the same set of wine types with exactly the same degree of detail, could represent their discriminatory conclusions in different ways: as chemical equations, points on a wine chart, identifying bundles of other properties (such as colour and viscosity) … or as particular taste sensations. And even if two such systems represent Burgundy by a certain taste sensation, Dretske has still given us no reason to commit ourselves to the claim that it will be the same taste sensation: the possibility remains that one could experience the taste of every red wine as being like what I experience when I sample a fine Burgundy, and the other could taste like what I experience when I sip a poor Chianti.34

What about the other argument, independent from his theoretical framework, that Dretske adduces (83–84) to support the identification of feeling, content and qualia? Here it is again:

i) qualia are supposed to be the way things seem in the sense modality in question;

34 This example also reiterates the possibility that the wine could taste like nothing at all: contrary to Dretske’s theory, it is possible some appropriately represented properties are not phenomenally experienced at all.
ii) things sometimes are the way they seem; therefore

iii) qualia are exactly the properties the object being perceived has when the perception is veridical.

The error here is perhaps best expressed as an equivocation over “the way things seem” in the two premises. Premise i) identifies qualia with the qualities of phenomenal mental experience—the taste of a strawberry, the sense-impression of the colour of a ripe apple. Premise ii) identifies mental content with facts about the external world—in effect, says that mental content is sometimes true. From this it does not follow that the qualities of phenomenal mental experience always correspond to the way the world would be if the experience were veridical. The premises of this argument would be true (let us suppose they are) and the conclusion false if strawberries ‘actually’ have taste $x$ but for some reason they have always had taste $y$ for person $S$, who can accurately pick strawberries out in the world just as well as anyone else, and for whom the taste $y$ happens to be how she represents the content “strawberry.” For $S$, strawberries taste like $y$, and usually when she thinks she tastes strawberry she really does taste strawberry; but “tasting like $y$” is not a property belonging to the strawberry. More radically, it might be that there is no ‘actual’ strawberry taste that strawberries always have, and which is exactly similar to what we experience, but that our taste sensations are how we represent some other complex strawberry property, such as chemical composition, to ourselves.

d) Conclusion

Dretske, then, has radically reconstructed qualia as non-mental, non-phenomenal properties, and has extended the class of qualia to include any property that can be discriminated by, at least, any evolved system with even a bare modicum of intelligence. One of Dretske’s main motivations for this redefinition, he makes plain, is the desire to render qualia ‘objective’ and accessible to study from a third-person perspective; however, there is little reason to believe that Dretske’s position is entailed by this desideratum (as I hope I have shown, a qualia–brain property identity thesis is at least as viable an avenue of research for example), and even on
Dretske’s account qualia turn out to be possessed of a regrettable slipperiness and possibly observer-relativity. Worse, Dretske’s theory falls victim to the good old fashioned problem of the conspicuous absence of qualia-holders during misperception—the argument from perceptual relativity.

Most centrally though (and in this he is far from unique), Dretske fails to capture the phenomenal kernel at the core of the notion of qualia. I suspect that the central reason for this absence at the heart of Dretske’s account of conscious mental experience is a conflation of qualia with content (where “content” is understood, at least on an intuitive level, as dealing with the semantic or representational, rather than with the phenomenal per se\(^{35}\)): that is, Dretske possibly feels that having given what is arguably a complete account of what mental states indicate, and of how they go about performing that function, he has completed the task of describing the contents of the mind. But this is not so; as I have suggested above, one and the same content can be represented in a number of different ways, some with different ‘feels,’ and some without any feel at all. Consider one last example: Dretske’s dogfish (81 ff.), mentioned above. Dretske is explicit in saying that there is no more to the content of a dogfish’s experience of an electric field than that the field “is normal here, it is ‘pinched in’ there, and it ‘bulges’ between here and there” (84). The only ‘extra’ element to the dogfish’s experience of the electric field, beyond this geometrical description, is that the fish is related to the field such that, if fish could think, it could in principle think to itself “that this (referencing to an experienced electric field) has pattern P” (87). Suppose that Dretske is correct; that he has succeeded in giving a full account of the content of a dogfish’s experiences of an electrical field. This very same content is also represented by the readings on the screen of an ammeter mapping the field; which, if it could think, might think “this field has pattern P.” But from this identity of content it does not follow that the dogfish and the ammeter experience the same qualia; in fact, the ammeter experiences, not just different qualia, but no qualia at all.

---

\(^{35}\) This, for example, is the way Robert Cummins defines mental content (1989, 11–12), stating that the central issue in this domain is “What is it for a mental state to have a semantic property?” By contrast, the central issue in dealing with qualia is “What is it for a mental state to have a phenomenal property?”
3. CHALMERS’ DUALISM

Chalmers’ 1996 book *The Conscious Mind* has recently sparked great interest by arguing that physicalism must be false: that phenomenal consciousness cannot possibly be reduced to any set of physical facts (e.g., in my terms, qualia are atomic properties), and that this is provable *a priori*. This makes Chalmers a rather different kind of opponent from the other two: Dennett and Dretske in effect deny the existence of qualia as I have described them; Chalmers presents a much more similar picture of qualia but denies that such properties are consistent with physicalism.\(^36\) His book is a careful and clever presentation of some of the strongest arguments available for property dualism, and in most points of detail it strikes me as very convincing: like many other commentators, though, I find that it has a basic mistake at its core which robs Chalmers’ dualist conclusions of their force.\(^37\)

\(\text{a) Chalmers’ Modal Argument}\)

Chalmers’ argument for the claim that phenomenal consciousness (I shall speak here of “qualia” from now on) is necessarily irreducible is straightforward enough. As he summarises it (1996a, 123) it goes:

i) In our world, there are conscious experiences.

ii) There is a logically possible world physically identical to ours, in which the positive facts about consciousness in our world do not hold.

iii) Therefore, facts about the consciousness are further facts about our world, over and

\(^{36}\) My own position, the reader will recall, is that physicalism remains unproved with respect to qualia but is still a live option—thus, not falsifiable *a priori*—and is indeed the most attractive option. This is how Chalmers’ position comes into conflict with that defended here. Once the basic mistake mentioned below is removed from Chalmers’ account, his conclusions seem to me to be transformed into propositions very much like my own—that qualia are physical if they supervene appropriately upon micro-physics; that properties which do not so supervene are highly bizarre; that the phenomenality of consciousness must be addressed seriously; that the first-person epistemology of qualia is interestingly affected by the phenomenal apprehension of qualia; and so on.

\(^{37}\) However my own account of just where Chalmers goes wrong is, I believe, rather different than the ‘standard’ story.
above the physical facts.

iv) So materialism is false.

That is, he says, “[i]f a physically identical zombie world is logically possible, it follows that the presence of consciousness is an extra fact about our world, not guaranteed by the physical facts alone” (123). Furthermore, physically identical zombie worlds are logically possible.

The most disputable claim here, obviously, is premise ii). Chalmers defends it in the pivotal chapter three of his book, “Can Consciousness be Reductively Explained?” It is also possible to object to the transition from premise ii) to sub-conclusion iii) on the basis of the claim that it is metaphysically possible worlds which are of interest, not logically possible ones, and Chalmers has not shown that the displacement of qualia is metaphysically possible. Chalmers addresses this complaint in chapter two section four, and chapter four section two. I shall describe Chalmers defence of each aspect of the argument in turn.

i) Qualia and “logical supervenience.”

What Chalmers means by “logically possible” in premise ii) is, he stipulates, “conceptually possible,” which in turn is to say, “conceivable on ideal rational reflection.” (The restriction to “the positive facts about consciousness” is intended to deal with the problem in defining physicalism, which we noticed above (Chapter Eight, section 2a), of the possibility of there being extra unphysical stuff.) Chalmers’ claim in premise ii), then, is that if we fix everything physical about our world we can still conceive of the possibility that qualia could be distributed differ-

---

38 He introduces logical possibility as “possibility in the broadest sense, corresponding roughly to conceivability, quite unconstrained by the laws of our world,” and says flying telephones are logically possible while male vixens are not (1996a, 35). He re-emphasises this point in his forthcoming “Materialism and the Metaphysics of Modality.”

39 Chalmers’ solution, in brief, is to formulate the supervenience claim as dealing only with the set of supervenient facts true of this world, and he is then forced to restrict this set of supervenient facts still further to include only “positive facts,” i.e. those which do not include or entail negative existential claims. This is because, clearly, it would be a problem for Chalmers’ unmodified world-specific formulation if fixing the physical facts also fixed such facts as that there are no angels or that all living things are based upon DNA. The reader may or may not consider this a much clunkier solution to the problem of “alien stuff” than my own, presented in Chapter Eight.
ently than they actually are.

He defends this claim in three different ways, presenting five arguments in their support. First, he argues directly that certain situations where qualia are redistributed over the physical are *conceivable*. Here he cites the logical possibility of zombies (94–99), and of inverted spectra (99–101). Second, he argues that knowledge of physical facts cannot entail the right kind of knowledge of phenomenal facts, and brings up the epistemic immediacy of qualia (101–103) and Jackson’s knowledge argument (103–104) in this connection. Finally, he suggests that there is available no analysis of the concept of phenomenal consciousness which might ground an entailment from the physical—basically, that functional and structural analyses of phenomenal consciousness are in principle inadequate (104–106, 111–121).

Each of these five arguments operates by making *vivid*, in different ways, the fact that we can imagine the physical remaining identical while the phenomenal varies. This itself, however, is not really in much dispute: what matters is what *implications* this psychological ability has—often, for example, it is objected against arguments of the sort Chalmers uses that “conceivability is an imperfect guide to possibility.” First, it might be said that conceivability has nothing to do with *logical possibility*: Chalmers, however, *defines* “logical possibility” as “conceivability on ideal, rational reflection,” and I think makes a reasonably good case for doing so (52 ff.)—at any rate, I won’t quibble with that move here. Second, it might be said that *de facto imaginability* has nothing to do with “conceivability on ideal, rational reflection.” This is the real sticking point for Chalmers, on which as far as I can see he has little to say: I shall return to this point in a moment. Third, it might be said that, as Kripke showed in 1971, *a priori* conceivability has nothing to say about *a posteriori* necessity and it is the latter kind of possibility which is of interest here; Chalmers thinks this point is central and considers it in detail.

ii) Qualia and “metaphysical supervenience.”

To establish that the crucial move between steps ii) and iii) of his argument goes through, Chalmers tries to show that logical possibility is, relevantly, *the same thing* as metaphysical necessity. If he can show this, then he can claim that all logically possible worlds are also metaphysi-
cally possible, and so, given premise ii), that there is a metaphysically possible world, physically identical to our own, in which qualia are differently distributed. This would demonstrate the falsity of the position of those Chalmers calls “type-B materialists” who accept that phenomenal facts are not necessitated *a priori* by physical facts but hold that that they *are* so necessitated *a posteriori*: that is, those who might agree with Chalmers that the non–physicalist-supervenience of qualia is *imaginable* but deny that it is therefore actual. After all, objectors of this type say, we couldn’t tell *a priori* that water is H₂O but it turns out that it necessarily is; perhaps, similarly, though we can’t tell that certain third-person observable complexes of brain properties are identical with qualia, nevertheless they are.

Chalmers defends his identification of the logically and the metaphysically necessary with respect to qualia in three stages. First he tries to show, on the basis of a “two-dimensional account” of *a posteriori* necessity, that the set of metaphysically possible worlds is exactly the set of conceivable worlds. Then he applies the 2-D account to an argument that *a posteriori* necessities of the usually cited variety cannot save physicalism. Finally he argues that we have no good reason to believe in other sorts of *a posteriori* necessity (which he calls “strong necessities”) which might escape the 2-D account.

Chalmer’s 2-D account of *a posteriori* possibility is designed to show that the distinction between metaphysical and logical possibility does not distinguish between sets of possible *worlds* but instead demarcates between evaluations of statements over the set of all possible worlds. Kripke’s work in *Naming and Necessity* (1980), according to Chalmers (57ff.), is best understood as providing a “two-dimensional” picture of meaning and necessity, in which (if we understand intensions as functions from possible worlds to referents) certain concepts turn out to have *two* intensions. The primary intension is the function from worlds to extensions reflecting the way that actual-world reference is fixed. Thus the primary intension of “water” is all the *watery-stuff* in actual and possible worlds; if the actual world turns out one way, a concept will pick out one

---

40 Type-A materialists, on Chalmers’ taxonomy, hold that qualia (if they exist at all) are necessitated *a priori*: this group are supposed to have been dealt with by Chalmers’ defence of premise ii) in his chapter three.
thing (e.g. H\textsubscript{2}O), but if it turns out another way it will pick out something else (e.g. XYZ). The secondary intension picks out water in counterfactual worlds given that it has turned out to be H\textsubscript{2}O and once the primary intension has been rigidified.\textsuperscript{41} Thus, the secondary intension of “water” picks out H\textsubscript{2}O in all possible worlds.\textsuperscript{42}

It is the primary intension, according to Chalmers, which is central with regards to the explanation and supervenience of natural phenomena: this intension is the one that captures what needs explaining. For example, in thinking about water what requires explanation is the watery-stuff in our environment, and the discovery that this stuff is H\textsubscript{2}O comes only after or during this explanation. Furthermore, primary intensions are independent of empirical factors: “the [primary] intension specifies how reference depends on the way the external world turns out, so it does not itself depend on the way the external world turns out” (57). To put it another way, we have to first fix reference in the actual world (the primary intension) before we can fix reference in counter-factual worlds (the secondary intension).

What this gives us, Chalmers suggests (62 ff.), is two kinds of meaning—\textit{a priori} and \textit{a posteriori} aspects of meaning respectively—and more importantly two kinds of logical necessity, which he identifies as corresponding to logical necessity and metaphysical necessity. Thus, using the primary intensions of “water” and “H\textsubscript{2}O,” it is logically possible that water is not H\textsubscript{2}O (since the intensions differ); using the secondary intensions, the intensions of “water” and “H\textsubscript{2}O” are the same so it is not metaphysically possible they differ. Matters of logical possibility are \textit{a priori}, at least in principle; matters of metaphysical possibility are generally accessible only \textit{a posteriori} since facts about the actual world may play a role in determining the secondary intensions.

However, according to Chalmers, \textit{both} of these kinds of necessity apply to \textit{statements} only, and not to worlds.

\textit{[T]he oft-cited distinction between ‘logical’ and ‘metaphysical’ possibility stemming

\textsuperscript{41} That is, “water” is “\textit{dhat} (watery-stuff)” (to use terminology from Kaplan 1979).

\textsuperscript{42} In short, reference in a counterfactual world depends both on that world and on the way the actual world turns out—as Chalmers puts it (61), certain concepts determine a doubly indexed function.
from the Kripkean cases—on which it is held to be logically possible but not metaphysically possible that water is XYZ—is not a distinction at the level of worlds but at most a distinction at the level of statements. A statement is ‘logically possible’ in this sense if it is true in some world when evaluated according to primary intensions; a statement is ‘metaphysically possible’ if it is true in some world when evaluated according to secondary intensions. The relevant space of worlds is the same in both cases.

(67–68)

The objection that Chalmers illicitly slides from logical to metaphysical possibility, on this story, comes down to the fact that Chalmers’ arguments work only for the primary intensions of the concepts involved and not for their secondary intensions (132). Thus, the objection runs, Chalmers might have established that the primary intension of “phenomenal redness” is different than that of any physical concept at all—that is, that we can imagine them coming apart—but he has yet to show they are not in fact identical: that phenomenal redness is not identical in every metaphysically possible world with some physical property.

Chalmers’ response to this objection is simply the following: what we wish to describe and explain is the property picked out by the primary intension of phenomenal terms.

If we can show that there are possible worlds that are physically identical to ours but in which the property introduced by the primary intension is lacking, then dualism will follow. … This difference in worlds is sufficient to show that there are properties of our world over and above the physical properties. By analogy, if we could show that there were worlds physically identical to ours in which there was no watery stuff, we would have established dualism about water just as well as if we had established that there were worlds physically identical to ours in which there was no H₂O. And importantly, the difference with respect to the primary intension can be established independently of a posteriori factors, so that considerations about a posteriori necessity are irrelevant. (132–133)

Thus, for example, nothing about Kripke’s a posteriori necessity renders any logically possible worlds impossible, it just tells us that they are misdescribed. For example, these are worlds where it is false to say that “water” (secondary intension) is XYZ (134).

Chalmers also argues that the primary and secondary intensions coincide for phenomenal terms: “if something feels like a conscious experience, even in some counterfactual world, it is a conscious experience. All it means to be a conscious experience, in any possible world, is to have a certain feel” (133). His argument, however, is designed to go through even without this claim.
Finally, Chalmers considers the possible view—which he calls *strong metaphysical necessity*—that, though Kripkean *a posteriori* necessity has no force against his argument, there may nevertheless be some other reasons to think that there is a ‘space’ of metaphysically possible worlds which is more constrained than that of the logically possible worlds: that is, some worlds are logically possible but not metaphysically possible. Chalmers’ answer, in short, is that there is no good reason to believe in such a modality. “Such ‘metaphysical necessities’ will put constraints on the space of possible worlds that are brute and inexplicable. It may be reasonable to countenance brute, inexplicable facts about our world, but the existence of such facts about the space of possible worlds would be quite bizarre” (137).

*b) Chalmers’ Mistake*

I am prepared to accept that Chalmers has established that “conceivability on ideal, rational reflection”—with a heavy emphasis, as we shall see, on the “ideal”—is a reliable guide to the kind of possibility relevant to his general argument for irreducibility. Furthermore I agree that if it is logically possible that one could fix the totality of the physical—including its laws—in some possible world and still vary the phenomenal, then we will have established that qualia are irreducible to the physical.45 Finally, I am persuaded that the necessity of this supervenience must be in some sense *a priori*, rather than *a posteriori*: physicalist reductionism would be true only if, given that we knew everything there is to know about the physical, we could have predicted the existence and disposition of qualia46 before apprehending them (and could presumably predict qualia in others without ever phenomenally apprehending them).

Nevertheless, despite all this common ground, I still think one is right to object to Chalmers that *imaginability is not a reliable guide to possibility*, and that this really is the rock upon which his

45 This is really just the result we established in Chapters Seven and Eight, with the supervenience claims expressed in terms of logical necessity within what I called the set of physically possible worlds: that is, given that the laws of physics and all the boundary conditions are what they are, qualia are physical only if they are fixed by logical necessity.

46 Given also, as Chalmers admits, that we already possess the necessary concepts of qualia (36, 70).
dualism founders. What is important is to properly focus the objection so that it comes in under Chalmers’ defences, and pin-points what is really wrong with his account, rather than merely apparently wrong with it.

The problem, then, is not that ideal a priori conceivability fails to establish modal claims; the problem is that the de facto abilities of today’s human beings to imagine things are insufficient to establish claims about ideal a priori conceivability. In a nutshell, that I can imagine an entity just like me except that he is a zombie is not enough to show that it really is conceivable, for the ideally rational agent with all the physical information, that such a zombie is possible. It is the latter kind of claim that Chalmers needs to establish premise ii) of his argument, but the evidence he gives is sufficient only to establish the former sort of claim: this is the gap in Chalmers’ toilings.

Chalmers makes one comment in his book which is directed towards objections of this general sort. He rejects the general claim that conceivability does not entail possibility in the following way:

When it comes to matters of explanation, it is clear that conceivability is central. If on reflection we find it conceivable that all these physical processes could take place in the absence of consciousness, then no reductive explanation of consciousness will be satisfactory: the further question of why we exist and not zombies will always arise. Even if conceivability is tied to the limits of human capacity, explanation is tied to the limits of human capacity in a similar way. (110)

However, with this (again) we can safely agree without committing ourselves to dualism: it is true, let us suppose, that some account is not properly explanatory for me if I can imagine that the facts the account describes are true but those which are to be explained are false or different. Firstly, however, there is more than one way in which an account can fail to explain and, as we saw in Chapter Nine section 2c, one of the central ones is that it might be too complicated

47 As we might put it, the objection is not Is conceivability a reliable guide to possibility?, but Is conceivability a reliable guide to conceivability?

48 I make basically this very point above, in Chapter Eight, section 2b, and my presentation here is designed to supplement and focus what I say there, but not cover exactly the same ground again.
or stuffed with irrelevancies for me to properly understand it.\textsuperscript{49} Thus, an account may be ‘genuinely explanatory’—in the sense, perhaps, that someone who did understand it fully would no longer be able to ask “but why did the explanandum occur?”—yet I might hear that account, even believe (wrongly) that I fully understand it, and still be able to imagine the explananda varying independently of the explanans.\textsuperscript{50}

Secondly, even if we do fully understand some putative ‘explanation’ and that account fails to ‘fix’ the phenomena it is supposed to explain, all this shows of course is that this is not a good explanation, not that no good explanations are in principle available for these phenomena. That this item in some class is not $P$ does nothing by itself to show that no members of that class are $P$. At best, then, Chalmers has established that none among the current crop of candidate ‘explanations’ is actually explanatory; he has not established that no physicalist explanation is available.

Chalmers has one more response to what he calls “the standard ‘how do we know we can really conceive this’” objection; the “real point,” he says, is “the challenge to isolate an inconsistency in the notion [of, e.g., zombies], and in particular to demonstrate the aspect of the concept of consciousness which could ground a conceptual entailment from the physical facts to the facts about consciousness” (1996b). That is, Chalmers clearly intends his examples—zombies, inverted spectra, etc.—to show not just that we can’t imagine that fixing the physical rules these things out, but that it is something like a conceptual truth that zombies are physically possible. How can this be? Well, perhaps it is because, as Chalmers is at pains to point out, for him logical necessities are discoverable “a priori”: thus, he might say, his challenge is for someone to find an a priori proof that zombies are physically impossible, and the absence of such a

\textsuperscript{49} As we saw above, not all higher level patterns are easily visible from the bottom level, though they may nevertheless exist there.

\textsuperscript{50} “Of course we can imagine ‘all these physical processes’ in a brain without involving consciousness, just as we can imagine ‘all these physical processes’ in a flying baseball without involving a flying baseball. I seriously doubt anyone—not even a philosopher with super-imagination—would suddenly conclude ‘flying baseball!’ if presented with a mountain of data about positions and momenta of every single particle in a baseball. Chalmers offers no reason to believe the two cases differ in a relevant way.” (Mulhauser 1996)
proof is itself very strong evidence that none is available. In other words, if zombies were physically impossible we should be able to prove it \textit{a priori}—on the basis of the primary inten-
sions of the concepts involved, as Chalmers would say. Since we cannot produce this proof, zombies must be physically possible … and, perhaps, it is the \textit{non-availability of such an a priori proof} that Chalmers really finds so intuitively compelling (rather than merely the blank ‘imagin-
ability’ of zombies).

But all of this cannot be quite right: for, notice the role that the term “\textit{a priori}” is actually playing in Chalmers’ account. His claim is not that we could have predicted the existence of qualia, or even the entailment of qualia by the physical, \textit{in utero}, so to speak. His claim is that \textit{given we know all there is to know about the physical fact-totality} we could predict the existence and disposition of qualia. Similarly, Chalmers does not assert that we could always have predicted that life was physically reducible\footnote{“Vitalism was mostly driven by doubt about whether physical mechanisms could perform all the complex \textit{functions} associated with life: adaptive behavior, reproduction, and the like” (1996a, 109).}: instead, he says, once we found out enough about the physical mechanisms involved we could \textit{then} see that these mechanisms necessitated, and so explained, the phenomenon of life. Reducible properties are predictable \textit{a priori} from sufficiently com-
plete knowledge of their physical bases; without such knowledge, they need have no \textit{a priori} connection to the physical at all. The moral of this is clear: as I have said before, modal argu-
ments of the type used by Chalmers \textit{presuppose} that we have adequately complete knowledge of the micro-physical subvenients for qualia, and there is currently \textit{no reason at all} to think our knowledge of the physics of qualia is complete. As Patricia Churchland puts it:

\begin{quote}

Perhaps we like to put our ignorance in a positive light, supposing that, but for the Pro-
fundity of the phenomenon, we \textit{would} have knowledge. But there are many reasons for not knowing, and the specialness of the phenomenon is, quite regularly, not the real reason. (1996, 406)

For this reason, Chalmers fails to establish the \textit{truth of} the claim that qualia are irreducible to the physical; all he has done is to demonstrate that they \textit{may} be so irreducible—that we do not at present know otherwise. One of the possible explanations for this uncertainty is that qualia
really are a uniquely non-physical kind of phenomenon; but another explanation, which has not yet been ruled out, is that we simply do not yet know enough about the nature of the dependence of qualia upon the brain—about the physical facts involved—to understand how they are reducible.
Chapter 12: Conclusion

In this concluding chapter I intend merely to draw together and briefly summarise my main conclusions concerning the status, metaphysics and epistemology of qualia, and then indicate some of the directions in which I think future philosophical research in this area would be fruitful.

1. SUMMARY

What I have tried to do in this dissertation is to develop, in enough detail that its promise can be reasonably assessed, a realist, internalist, physicalist account of qualia.\(^1\) I have attempted to provide \textit{a priori} compelling reasons to believe that qualia are real and internal; the question of physicalism, by contrast, is an \textit{a posteriori} matter for which presently the evidence either way is inadequate, and so I have contented myself merely with arguing that physicalist accounts of qualia are \textit{possible} (i.e. not already known to be false) and, more than that, \textit{attractive} in their promise of theoretical fruitfulness.\(^2\)

I began by defining “qualia” simply as “properties as they appear in experience” where the relevant contrast is to something like properties as they are ‘in themselves’ if the latter are different properties. Thus the quale redness is redness as it is ‘presented to us’ in visual experience, and (as far as the meanings of the words are concerned) it may or may not be different than actual redness (which is a property of external objects). However, though it is not an analytic truth that qualia are internal psychological properties—loosely speaking, ‘properties of conscious-

\(^1\) Obviously, at least in my terminology, any account of qualia is also effectively an account of experiences. Furthermore, since I hold that phenomenal consciousness can be completely characterised as, roughly, a seamless sequence of qualia-tokens which can then be carved up into experiences (rather than being, for example, an extra ‘thing’ over and above our qualia), I also take myself to have provided a fairly full account of phenomenal consciousness. However, I have not argued explicitly for this latter claim in this work. (My paper “Beyond the Fringe,” forthcoming, contains some preliminary thoughts on this issue.)

\(^2\) I also suggested in, for example, my discussion of atomicity that physicalist accounts are antecedently more \textit{plausible} than their alternative.
ness’—I argued that it is an a priori truth. I made this claim on the basis of a reformulation of the arguments used most recently by the sense-datum theorists of the first half of this century, and in particular arguments from perceiver relativity. My reformulation was designed to avoid the various problems in the arguments which led to the demise of the sense-datum; I argued at length that my version of the argument from illusion is sound and establishes that properties as they appear in experience—qualia—are at least sometimes distinct from properties of externally perceived objects. Furthermore, since every property instantiation is a property of some individual, “there must be something which is phenomenally grey when the colour-blind perceiver looks at a ripe strawberry ... and it isn’t the strawberry.”

The pressing question then became: in what do qualia inhere? What is phenomenally grey if it is not the strawberry? I argued that, barring a return to substance dualism, the only plausible candidates for these “qualifiers” are (at least in humans and other animals) brain states, and then tried to press home the significance of this conclusion: these brain states do not just have the property of indicating that some external object is such-and-such a colour—this argument shows that they actually instantiate phenomenal versions of these colours, they actually are phenomenally red, phenomenally painful, and so on. Now since brain states are obviously not actually red—they do not have the same colour property as ripe strawberries and London buses—it is plain that we cannot model phenomenal colour on ‘actual’ colour. Instead, I argued, it can only be that to token phenomenal property $F$ is to be the phenomenal sensation of $F$. Thus, phenomenally red brain states are not actually red, don’t ‘look red’—they have the property of being the sensation of red. I noted, however, that this by itself is a neutral claim with respect to mind-body identity, and does not automatically falsify physicalism.

From this firm base—that qualia are real, internal, full-bloodedly phenomenal, and possibly physical—I embarked on an examination of the epistemology of qualia, asking primarily whether and in what sense qualia are “given.” I distinguished between the phenomenal apprehension of qualia—which is just what it is to have qualia—and propositional apprehension of qualia, which involves having beliefs about one’s own qualia. I then analysed givenness into three components—immediacy, certainty and privacy—and discussed each in turn. I concluded
that phenomenal apprehension of qualia is indeed in some sense “immediate,” that it is “certain” in the sense of being indubitable, evident and incorrigible, and that it is de facto highly “private.” However, I tried to show that, although qualia do turn out to be “given” after all, this adds nothing problematic either to the notion of qualia or to the prospect of physicalism: to be “given,” it turns out, is not necessarily to be especially mysterious or incoherent.

On the other hand, I agreed that qualia are not “Given” in the very strong sense accepted by the logical positivists but rejected by Sellars and Wittgenstein, that of providing an absolutely certain epistemological foundation—a set of indisputable propositions about facts. Propositional knowledge of our own qualia, I showed, can (rather simplistically) be said to be generally more “certain” than propositional knowledge of the “external world,” but has no absolute certainty. Finally, though, I raised the prospect—ignored or rejected by Sellars and Wittgenstein—of phenomenal apprehension providing some kind of check upon our cognitive attitudes towards our own qualia.

The next question I addressed was the question of what can be known a priori about the metaphysics of qualia. My starting point was that “qualia are not concepts, relations or logical entities—they are properties of concrete physical objects; and the best way to find out about particular properties, prima facie, is to go out into the world and study them.” Nevertheless, I set out to see what could usefully be said from the armchair about the relation between qualia and the brain (in addition to the fact that they are properties of the brain), and about whether qualia are consistent with the truth of physicalism.

I began with a unifying analysis of the family of supervenience relations, and showed that supervenience is not a ‘new’ kind of dependence relation but instead a set of determination relations; supervenience claims are therefore descriptive rather than explanatory. I then applied this analysis to the case of qualia and used my schema for supervenience to define what it would be for physicalism to be true with respect to qualia: roughly, qualia are consistent with physicalism iff fixing the physical fixes a certain set of qualia in every physically possible world.

---

3 Or at least nothing more problematic than the basic fact that they are phenomenal properties.
and the complete set of qualia in the actual world. The question whether physicalism actually is true for qualia is not decidable \textit{a priori}, however: I supported this claim by demonstrating that the main arguments for and against physicalism—the logical possibility of zombies and the impossibility of epiphenomenalism—fail. We shall only discover whether qualia are consistent with physicalism, I suggested, when we discover in more detail how qualia depend upon the brain: whether, for example, they are identical with, or constituted by, complexes of objectively observable brain properties.

The actual nature of qualia’s dependence upon the brain is something that (contra Chalmers) requires empirical, not directly philosophical, investigation. However there are some relevant conceptual issues that await philosophical clarification, and among the most important of these, I suggested, are claims to the effect that qualia are “higher level” or “emergent” properties in the sense that they are \textit{ontologically irreducible} to neurones, but nevertheless perfectly real and potentially physical properties. I conducted my discussion of this issue by first making a taxonomical analysis of the notion of “being at a higher level than,” and then applying this analysis to qualia. I argued most centrally that, for qualia to be ontologically irreducible in any interesting way their low-level ‘realisations’ would have to conform to no pattern at all—would have to have nothing at all uniquely in common. This, though certainly possible, would be an extremely, and perhaps unexpectedly, radical claim, and in one of its versions would apparently violate even minimal supervenience claims for qualia. I also argued that qualia are probably neither “ontologically atomic” nor “ontologically basic” properties.

Finally, I considered three alternative accounts of qualia and argued that Dennett, representing the eliminativists, makes no arguments which impact upon the kind of theory of qualia defended here; that Dretske, standing in for the externalists, fails to deal with the argument from perceiver relativity and confuses phenomenal properties with perceptual content; and that Chalmers, a leading property dualist, builds an elegant house of cards upon the false premise that \textit{de facto} imaginability is a reliable guide to ideal conceivableability.
2. PROBLEMS REMAINING FOR ANOTHER DAY

I shall end by briefly mentioning some important and attractive avenues of research that seem to me to proceed onwards from the point at which we have arrived. Most pressingly, of course, it is vital that the dependence relation between qualia and the brain be further studied: what it is most important to discover now is what sort of brain property are qualia, and how are they related to the third-person observable brain properties. This, as I have emphasised, is primarily an empirical and not a philosophical matter, but this is also the area in which the single biggest conceptual problem arises: the “mystery of consciousness,” it seems to me, comes down to the question: What is it to be a phenomenal property? How, as I put it above, does phenomenal-ness ‘work’? How can it be that brain states can instance a token of that very property which is ‘occurrent’ when I look at a blank red wall? I have made virtually no progress at all towards answering this question—Chalmers’ “hard question”—in this dissertation, but that of course is not to say that no philosopher ever could substantially assist the empirical sciences in approaching a solution. (My present concern has been slightly different: to show that the mere existence of this “hard problem” does not in itself falsify physicalism; and to show that relatively extensive and fruitful philosophical theorising about qualia is possible despite the intransigence of their phenomenal nature.)

Second, and perhaps almost as mysteriously, there is what we might call the qualia binding problem. This is neither clearly an empirical nor a clearly philosophical problem. Experiences, I have suggested, are structured sets of qualia tokens; these qualia tokens, presumably, may be scattered around various regions of the brain and may occur at slightly different times—indeed, I suppose it is possible that each individual quale is distributed in this manner. How can it be, then, that these property tokens ‘arrange themselves’ so that from the first person we experience the kind of seamless phenomenal manifold that we do?

---

4 Indeed, it seems to me and I have tried to suggest throughout that theorising about qualia is much more fruitful—as well as more honest—for those who recognise that qualia really are phenomenal.

5 I wonder if this problem is somewhat alleviated—in comparison with the traditional “binding problem,” for example—by the fact that qualia, and only qualia, are phenomenally apprehended. They therefore form some-
Thirdly, there are a group of pressing philosophical issues to do with the relation between qualia, what I have been calling attention, and mental content. The field here seems to me to be vast and rich. For example, how do our conclusions about qualia connect with theories of perception? I take it that we do not phenomenally apprehend our qualia and then somehow make inferences from this ‘data’ to objects in the external world—instead, qualia are most plausibly properties of certain stages of the more-or-less inferential, information-processing, human perceptual system [... insert your favourite theory here]. Must they, then, be by-products of this process, or is there a manner in which the phenomenal properties could play a role in this system? What is the relation between our beliefs about our own qualia and our beliefs about objects and events in the external world?

Another interesting question in this domain, already touched on above: assuming that qualia are not themselves knowledge of facts, and assuming with Sellars and Wittgenstein that the kind of factual knowledge that can feature in the normative game of epistemic justification is necessarily linguistic and propositional, is there nevertheless an intermediate form of apprehension which can be called “perceiving that”? Is there any sense at all that, in experiencing qualia, animals and babies acquire knowledge? Relatedly, just how strong a constraint do qualia exercise upon our beliefs: under what conditions, if any, could we have beliefs about the external world or about our own qualia which blatantly conflict with our qualia?

Fourth, there are several metaphysical questions about reducibility and identity that I have not addressed here and which have clear relevance to the problem of qualia. It would certainly be relevant to show, for example, that phenomenal properties are multiply realizable in such a way that they cannot be type identical with brain properties under any level of description, but merely, perhaps, realised by those properties.  

thing of a self-selecting class from the first-person perspective—there is no need to additionally ‘code’ for being a quale and then ‘bring together’ these coded items. Furthermore, one might speculate that phenomenal colours, shapes, smells etc., are in some way self-arranging as well—after all, one might say, how else could a colour be combined with a shape except as a coloured shape? These, of course, are the merest first inklings (if that) of how to go about thinking about this problem.

6 I suspect, in fact, that this is not so—that qualia are multiply realisable in no more exciting a way than, say, being
Finally, there are the vestiges of a private language argument to be dealt with—or, to put it another way, the problem of developing an effective vocabulary for describing qualia. I have been speaking of “phenomenal redness” as if it were, basically, that kind of quale which is tokened when normal people see red things: however, this cannot be quite right (as a definition) because, in the absence of a completed science of qualia, I cannot be sure that the quale you token when you look at a ripe strawberry is of the same type as that which I token. If it were not of the same type—if it was a different colour sensation—then we certainly would not want to call them both “phenomenal redness.” The problem then arises: what are we to call them? Or, to put it another way, what do I really mean when I talk about “phenomenal redness”—what are the conditions for the correct application of that term?

\textit{watery-stuff} is multiply realisable, and that therefore they may perfectly well be identical with higher level third-person physical properties (which in turn are reducible to complex low level patterns)—but I have not made that argument here.
Bibliography


Bibliography


