Phenomenal consciousness

A naturalistic theory

Peter Carruthers

Professor of Philosophy and Director, Hang Seng Centre for Cognitive Studies, University of Sheffield

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011–4211, USA http://www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

First published 2000

Printed in the United Kingdom at the University Press, Cambridge

Typeface Monotype Plantin 10/12pt *System* QuarkXPress™ [SE]

A catalogue record for this book is available from the British Library

Library of Congress cataloguing in publication data

Carruthers, Peter, 1952–
Phenomenal consciousness: a naturalistic theory / Peter Carruthers.
p. cm.
Includes bibliographical references and indexes.
ISBN 0 521 78173 6 (hardback)
1. Consciousness. 2. Naturalism. I. Title.

B808.9 C37 2000
126 – dc21 99-086451

ISBN 0 521 78173 6 hardback

Contents

Figures

1 Assumptions, distinctions, and a map

The nature and aims of my project have already been explained in the Preface. In this opening chapter I shall lay out some of my background assumptions, introduce a number of important distinctions, and outline the direction which the discussions of later chapters will follow.

1 Physicalism and naturalism

In this section I shall briefly explain and defend two default assumptions, which form the background to the problem of phenomenal consciousness. It is these assumptions which appear to be challenged by the very existence of phenomenal consciousness, as we shall see in chapters 2 and 3.

1.1 Physicalism

One assumption I shall make is that we should at least *try* to be token-physicalists about the mind. We should maintain that all particular (or 'token') mental states and events are at the same time physical (presumably neurophysiological) states and events, if we can do so consistently with our other beliefs. In the present section I shall briefly motivate this assumption, which is shared by almost everyone now working in the philosophy of mind – which is not to say that physicalism itself is mandatory, of course; indeed, many of the arguments against physicalism derive from considerations to do with phenomenal consciousness, as we shall see.[1]

[1] There are many who would deny the claim that mental states and events are *neurophysiological* states and events, not because they reject physicalism, but because they endorse an *externalist* account of the individuation-conditions of mental states with intentional content, such as beliefs and desires (e.g. Burge, 1979, 1986a, 1986b; McDowell, 1986, 1994). On such accounts, the identity of a mental state is tied up with the identity and existence of the worldly objects and properties which that state is *about*. I shall ignore such views here, for simplicity only. The basic argument for physicalism can still go through, only with the complication that the mental cause of a bodily movement is a complex relational entity, involving both the brain state which is the immediate physical cause of the movement and relations to the items in the world which that mental state is about. The distinction between externalist and internalist (or between wide and narrow) accounts of intentional content will become important in chapters 3 and 4.

Descartes famously held that the mind is non-physical while the brain is physical, and that they interact causally with one another. For example, sensory stimulation causes conscious experience, and decisions cause bodily movements. One of the main objections to dualism ever since has been the difficulty of making sense of such a causal connection. Not that there is any problem of *principle* in understanding causal connections between physical and non-physical, in my view. For there is nothing in the concept of causation, as such, which requires all causes to be mediated by physical mechanisms. The real problem is to understand *how* such causation can occur, given what we already know about the physical world, and about causation in the brain.

Consider, first, the physical world in general. Most scientists now believe that physics is *closed*, in the sense of permitting no interference from, or causation by, events at higher levels of description (e.g. chemical or biological). On this view all atomic and sub-atomic events happen in accordance with physical laws (albeit probabilistic ones), and all events at higher, more abstract, levels of description must be realised in, or constituted by, those physical processes, in such a way as to allow no independent point of causal leverage. So while there may be chemical and biological laws, the events which figure in these laws must always, at the same time, fall under the laws of physics. On this picture there is simply no *room* for a distinct and independent psychological level, whose events are not physically constituted, but which can have an impact upon the physical behaviour of the body.

Consider, now, what is known about the brain. There is much still to learn, no doubt – about the functions and interactions of its parts, for example. But much is already known. It is known that the brain consists of nerve cells, of various known types. And much is known about how such cells function, and the physical causes which lead to their activity. Certainly there would appear to be no 'inverse causal black-holes' in the brain, such as would seem to be required by the interactionist picture (that is, there are no places from which brain activity emerges *for no physical reason*). Indeed, enough is already known about the brain to justify the claim that each event in the brain has a sufficient physical cause. So, again, the moral would appear to be that there is no room, here, for mental events to cause physical ones, unless those mental events are themselves physically constituted – that is to say, unless physicalism is true.

What are the alternatives to physicalism? One possibility would be to go for some sort of *panpsychism* (Nagel, 1979), believing that current descriptions of physical reality are inadequate, and that all physical events are in some sense already mental ones, or possess mental properties.

Another possibility would be to exploit the indeterminacies left open by physical theory at the sub-atomic level, to find a place for mental–physical interaction. It could be maintained, for example, that the mind somehow resolves all the sub-atomic indeterminacies which exist within the neurological events in our brains in one direction or another, thereby having an influence upon the overall patterns of activity in the brain (Penrose, 1994). Yet another alternative would be to embrace *epiphenomenalism* about the mental in general, or about phenomenal consciousness in particular, believing that conscious experiences are not physically constituted, and that while being caused by physical events in our brains, they can have no further physical effects in their turn (Jackson, 1982).

None of these alternatives to physicalism is at all attractive. For example, in connection with the last, there are real problems in explaining how we can know that we ourselves are phenomenally conscious, at least if it is allowed that intentional mental states like beliefs have a physical constitution (as does Chalmers, 1996). For then, by hypothesis, our belief that we enjoy experiences with *feel* will not be a product of those experiences themselves (but rather, at best, will be caused by the physical events which themselves cause such experiences), and would have occurred just the same even if brain events had *not* caused experiences.[2]

Alternatively, if the thesis of non-physicality is extended to intentional mental events as well as to phenomenally conscious ones, then our problem is to explain our knowledge of the mental states of others. For while our belief that we ourselves are phenomenally conscious may be caused by the presence of conscious experience, the mental states of others (and of ourselves) can have no causal impact upon behaviour. So even while someone is describing in technicolour detail how it feels to them to be undergoing a certain sort of experience, their behaviour provides no real evidence of the presence of such experience; for by hypothesis, they would have behaved just the same even if brain events hadn't given rise to mental events at all.

This is not the place to develop these and other objections to the various alternatives to physicalism in any detail. For my purpose here is just to remind the reader of the considerations which make physicalism the default option in the philosophy of mind. Unless there are very powerful arguments to the contrary, we should believe that all mental states and events are physically constituted. Most philosophers think that the strongest challenge in this regard is provided by phenomenal consciousness

[2] See section 3 below and chapter 5:3 for discussion of the distinction between intentional states such as beliefs and thoughts, on the one hand, and experiences on the other.

itself. Just how powerful this challenge really is will be considered in chapters 2 and 3, where it will be suggested that all the main anti-physicalist arguments commit fallacies of one sort or another.

1.2 Naturalism

Naturalism is the belief that all of the events and processes which occur in the world are natural ones, happening in accordance with causal laws. So there are no miracles, and everything which happens can in principle be provided with a causal explanation, or is subsumable under laws (albeit probabilistic ones). In addition, naturalism is normally construed as involving the idea that the different levels of causation in nature are *ordered*, in such a way that processes at higher levels are always realised in, and reductively explicable in terms of, those at the lower levels. This need not mean that all *properties*, or *types* of phenomena, are identical with types identifiable in terms of physics, since higher-level types (e.g. wings) may be multiply-realised in lower-level processes or structures (as in the differences between the wings of birds and the wings of bats). But it does mean that all higher-level properties should be physically *constituted*, in such a way that each instantiation of such a high-level property admits of reductive explanation into lower-level (ultimately physical) terms.[3]

These have been the guiding methodological assumptions of science. When puzzling events occur, scientists do not just accept them, and postulate a miracle. Rather they continue to probe and investigate, working on the assumption that there must be a causal explanation, if only they could discover it. And when scientists discover laws and law-like relationships in nature, they do not rest content with a heterogeneity of such laws. Rather, they assume that nature constitutes a *unity*, and they seek to understand the operations of some laws in terms of others.

Since these naturalistic assumptions have received ample – albeit not conclusive – vindication through the advancement of science, it should require some powerful considerations to overturn them in connection

[3] Not *everyone* accepts that naturalism must involve a commitment to the reductive explicability of higher-level phenomena into lower-level terms. Thus Chalmers (1996), for example, describes his dualist account of consciousness as 'naturalistic' – since he believes that the properties involved in consciousness are subject to natural law, and are linked with brain-events by basic natural laws – although he thinks that phenomenal consciousness cannot be reductively explained. However, this position, if correct, would be highly revisionary of our scientific world-view. The conception of nature as *unified* – in a way that requires commitment to the possibility of reductive explanation – is so deeply built into scientific methodology that it surely deserves incorporation into our understanding of naturalism. At any rate, this is what I shall assume in what follows (nothing substantive hangs on it – the point is merely terminological).

with the mind and mental phenomena. Our default assumption should therefore be that all mental events occur in accordance with causal laws, and that we may hope to explain both the operation of, and the properties involved in, those laws in lower-level (ultimately physical) terms. However, precisely what naturalism commits us to is important to get right; especially since the project of this book is to naturalise phenomenal consciousness. I shall return to the issue in more detail in the chapters which follow, especially in chapter 4. (But even in chapter 4 my discussion will be relatively superficial, digging just deep enough into the issues to serve my own explanatory purposes. For a much more extensive and sophisticated treatment, see Papineau, 1993.)

2 Functionalism and theory-theory

The assumptions in section 1 above relate to the metaphysics of the mind. In this section I shall say something about how I take the mind to be conceptualised, or conceived of. I shall be assuming that some form of functionalism provides the best account of the way in which we conceptualise mental states. Again the position is not entirely mandatory, and again some of the main challenges come from considerations having to do with phenomenal consciousness, as we shall see. But the advantages of functionalism as an account of the mind (*viz.* its metaphysical neutrality – hence allowing interactive dualism to be a conceptual possibility – and its solution to the problem of other minds) mean that it should not be given up lightly.

2.1 From Cartesian concepts to analytic functionalism

As I have just noted, the thesis under discussion in section 1 was metaphysical – it concerned what mental states themselves really *are*. But what of our mental-state *concepts*? Even if mental states turn out to be physical, that does not seem to be how we conceptualise them – Cartesian dualism is a conceptual possibility, at least, even if it is actually false.

The thesis often attributed to Descartes is that mental-state concepts are (at least at bottom) bare recognitional capacities – capacities to recognise the distinctive *feel* which our mental states have. More recently, Goldman (1993) has defended a version of this view – claiming that we know of our own mental states by direct recognition, attributing the feelings in question to others by a generally-reliable process of *simulation*.[4]

[4] For an extended critique of the simulationist position, see my 1996b, and Botterill and Carruthers, 1999, ch. 4.

While I shall accept (indeed urge) that *some* of our mental-state concepts are Cartesian in this sense – consisting in bare recognitional capacities for the subjective *feels* of experience – I shall argue in chapter 9:3 that such concepts are parasitic upon those which are more theoretically embedded. And there are a number of powerful arguments against any attempt to extend the Cartesian view to *all* mental-state concepts. The main ones are as follows:

(1) The Cartesian view makes it difficult to see how the idea of *non-conscious* mental states – or states which would lack any distinctive subjective feel – is even so much as a conceptual possibility. (See section 3 below, and also chapter 6.)

(2) There are many conscious mental states which seem to lack distinctive feels – for example, beliefs and abstract (as opposed to bodily) desires. Perhaps it may be replied that these states are *dispositions* – dispositions to engage in acts of *thinking*, which have felt properties. But even if (many) acts of thinking do have felt properties (by figuring in 'inner speech', say), they do not seem to be *conceptualised in terms of* those properties. And the idea of 'purely propositional' (unfelt) thinking does seem to be a conceptual possibility; indeed many people believe it to be actual.[5]

(3) The Cartesian view makes it difficult to see how we could ever acquire the rich causal knowledge which we manifestly do have concerning the operations of minds. Compare sense-data theory as an account of vision, which is the idea that we begin with capacities to recognise unstructured sense-data (such as colours and textures) and then build up to a complex causal representation of the world by a process of learning. No one thinks that this is a viable developmental story any longer in the case of vision; nor should they in the case of our common-sense understanding of the mind.

I shall assume, therefore, that the Cartesian view of mental-state concepts, when put forward as the basis of all mental-state understanding or

[5] See the results of Hurlburt's (1990, 1993) introspection-sampling studies. Subjects wore a modified paging device through the day, which delivered a beep via an ear-phone at irregular intervals. Subjects were instructed to 'freeze' the contents of their conscious awareness at the moment of the beep, and to make brief notes to be reported to the experimenters later. All normal (as opposed to schizophrenic) subjects reported instances of 'inner speech', in varying proportions; with most also reporting visual images and emotional feelings. Many also reported the occurrence of 'purely propositional' (wordless) thoughts. In my 1998c I argue that there may actually be no such thing as purely propositional conscious thought, and that these reports may really be the result of swift self-interpretation. But the argument is empirical, not conceptual. There seems no doubt that the *idea* of purely propositional thinking makes perfectly good sense.

as applying to all such concepts, should be rejected in favour of some alternative.

Most philosophers of mind over recent decades have claimed that we conceptualise mental states in terms of their distinctive causal roles, or *functions* (Lewis, 1966; Putnam, 1967; Stich, 1983). So for example, beliefs are states which are caused either by perception or inference or testimony, and which in turn interact with desires to generate intentions and actions. Pains are states which are caused by bodily damage or disturbance, which in turn cause the subject to have a desire to cry out, rub the offending part, and so on.

On this account, there is no problem in allowing for non-conscious as well as conscious mental states, provided that the difference between the two can be accounted for in terms of causal role. Nor is there any problem in allowing for mental states which lack feels. Moreover, it remains explicable that metaphysical dualism should ever have seemed an option. For although we conceptualise mental states in terms of causal roles, it can be a contingent matter what actually *occupies* those causal roles; and it was a conceptual possibility that the role-occupiers might have been some sort of *soul-stuff*. However, there are two main problems with analytical functionalism:

(1) It is committed to the analytic–synthetic distinction, which many philosophers think (after Quine's 'Two dogmas of empiricism' – 1951) to be unviable. And it is certainly hard to decide quite which truisms concerning the causal role of a mental state should count as analytic, rather than just obviously true.

(2) Some mental states seem to be conceptualised purely in terms of subjective feel, or with beliefs about causal role taking a secondary position, at least. For example, it seems to be the feel of pain which is essential to it (Kripke, 1972). We seem to be able to imagine pains which occupy some other causal role; and we can conceive of states having the causal role of pain which are not pains (which lack subjective feel).

These problems seem sufficient to motivate rejection of analytic functionalism, in favour of the so-called 'theory-theory'.

2.2 Theory-theory

A better variant on functionalism about mental-state concepts is to say that such concepts (like theoretical concepts in science) get part of their life and sense from their position in a substantive *theory* of the causal structure and functioning of the mind. (The other part they get from their

causal–referential relations to the items which they concern.) On this view, to know what a belief is (to grasp the concept of belief) is to know sufficiently much of the theory of mind within which that concept is embedded. All the benefits of analytic functionalism are preserved. But there need be no commitment to the viability of an analytic–synthetic distinction, if only because of the indeterminacy of 'sufficiently much'.

(Some of us also believe that our theory of mind – generally called 'folk-psychology' – is largely implicit and substantially innate, emerging in normal human children by means of maturation-in-a-normal-environment, rather than by a process of learning. It is very hard indeed to see how the theory could be acquired so early – by the age of three or four – by ordinary learning; and there are just the same patterns of genetically-caused breakdown which one would expect if it *were* innate – i.e. autism, widely thought to be a kind of *mind-blindness*. See Fodor, 1987, 1992; Leslie, 1991, 1994b; Carruthers, 1992a, 1996c; Happé, 1994; Baron-Cohen, 1995; Botterill and Carruthers, 1999, chs.3–4; Hughes and Plomin, 2000.)

What of the point that some mental states seem to be conceptualised purely or primarily in terms of feel? A theory-theorist can allow that we have *recognitional capacities* for some of the theoretical entities characterised by the theory. (Compare the diagnostician who can recognise a cancer – immediately and without inference – in the blur of an X-ray photograph.) But it can be claimed that the concepts employed in such capacities are also partly characterised by their place in the theory – it is a *recognitional* application of a *theoretical* concept. Moreover, once someone possesses a recognitional concept, there can be nothing to stop them prizing it apart from its surrounding beliefs and theories, to form a concept which is *barely* recognitional. Our hypothesis can be that this is what takes place when people say that it is conceptually possible that there should be pains with quite different causal roles.[6]

The only real competitors to a theory-theory account of our folk-psychological concepts are some combination of Cartesianism with simulation, on the one hand, or some sort of *interpretationalism* or *quasi-behaviourism*, on the other, of the sort defended by Davidson (1970, 1974, 1975), Dennett (1971, 1981, 1987), Gordon (1995) and various Wittgensteinians.

The former position has already been criticised briefly above. Here just let me mention, in addition, that since this position takes phenomenal consciousness for granted, its adoption would cut us off from the possibility of reductively explaining such forms of consciousness in cognitive

[6] It will be a consequence of the position to be defended in chapter 9:3 that purely recognitional (or 'Cartesian') concepts of experience, while perfectly possible, are actually parasitic upon a theoretical understanding of the subjectivity of experience.

terms – for if our concepts of the cognitive are grounded in awareness of *feel*, then we cannot use the former in reductively explaining the latter. The fruitfulness of cognitivist approaches to the problem of phenomenal consciousness – to be defended at length in this book – will therefore be a further, if somewhat back-handed, argument against the introspectionist–simulationist position.

As for interpretationalism, I believe that this sort of view is unacceptable in its *anti-realism*, failing to do justice to the realistic commitments of the folk (see Fodor, 1987, ch. 1; Davies, 1991; Botterill and Carruthers, 1999, ch. 2). So I shall assume that theory-theory is the default position to adopt, unless considerations to do with phenomenal consciousness can convince us otherwise.

3 Some distinctions: kinds of consciousness

There are a number of different notions of consciousness and/or a number of different kinds of use of the term 'conscious' which need to be distinguished carefully from one another. Failure to draw the right distinctions, and/or failure to keep the different notions apart, has vitiated much work in the area. What follows draws heavily on the work of Rosenthal (1986), Block (1995) and Lycan (1996).

3.1 Creature-consciousness 1 – intransitive

Sometimes we treat consciousness as an intransitive, non-relational, property of a creature. Here the *subject* of consciousness is the person (or animal); and consciousness is treated as a simple property of that person. So we speak of someone 'losing consciousness' and 'regaining consciousness'; we say of the coma-victim that he has not been conscious since his accident; we say 'I want you to make sure that my cat is not conscious during the operation'; and 'I was conscious all the while'; and so on.

Here 'conscious' seems to be more-or-less equivalent to 'awake'. Roughly speaking, to say of an organism that it is conscious (intransitive) is just to say that it is awake, as opposed to asleep or comatose. At any rate, it seems to be a *sufficient* condition for a creature to count as conscious at *t*, that the creature should be awake at *t*. It is perhaps more debatable whether wakefulness is also a *necessary* condition of intransitive creature-consciousness. For we might wonder whether or not we should say that people are conscious during periods of dreaming, even though we are quite clear that they remain *asleep* during dreams.

I suspect that what may be going on here is that we think that being a subject of conscious mental states – state-consciousness; see sections 3.4

to 3.6 below – is a sufficient condition for intransitive creature-consciousness; and what we are wondering is whether dream-experiences should count as conscious ones. I am more inclined, myself, to say that state-consciousness need not imply creature-consciousness. So I would be inclined to say that the dreaming subject is not conscious (hence requiring that wakefulness be necessary *and* sufficient for creature-consciousness), although the dreamer may be undergoing mental states which are conscious. But this point will not matter to what follows.

There does not seem to be anything especially philosophically problematic about intransitive creature-consciousness, as such. At any rate, the awake–asleep distinction, while no doubt interesting, does not seem to hold any particular difficulties for physicalist and theory-theory conceptions of the mental. And in so far as there *is* anything problematic about this form of consciousness, the problems derive from its putative conceptual connections with *state-consciousness*. The latter notion will be discussed below.

3.2 Creature-consciousness 2 – transitive

Besides saying of an organism that it is conscious (*simpliciter*) we also say of it that it is conscious *of such-and-such* (transitive), or aware of such-and-such. To say this is normally to say at least that the organism is *perceiving* such-and-such. So we say of the mouse that it is conscious of the cat outside its hole in explaining why it does not come out; meaning that it *perceives* the cat's presence. To provide an account of transitive creature-consciousness would thus be to attempt a theory of perception. No doubt there are many philosophical problems lurking here; but I propose to proceed as if I had the solution to them.

Two points about perception are worth making in this context, however. The first is that perceptual contents can be (and often are, to some degree) *non-conceptual*. While perception often presents us with a world of objects categorised into kinds (tables, chairs, cats, and people, for example) sometimes it can – and in the case of young children and many species of animal, presumably it often does – present a world which is largely unconceptualised, but rather presented as *regions of filled space* (Peacocke, 1992). Perception presents us with a complex array of surfaces and filled spaces, even when we have no idea *what* we are perceiving, and/or have no concepts appropriate to what we perceive. Imagine a hunter–gatherer transported to some high-tech scientific laboratory, for example – she may have literally no idea what anything that she is seeing *is*; but for all that she will see the distribution of surfaces, shapes and masses; she will have an idea which are distinct objects; which are liftable; and so on.

The second – related – point is that perceptual contents are *analog* as opposed to digital, at least in relation to the concepts we possess. Thus perceptions of colour, for example, allow us to make an indefinite number of fine-grained discriminations, which far outstrip our powers of categorisation and description. I perceive just *this* shade of red, with just *this* illumination, for instance, which I am incapable of describing in other terms than, 'The shade of *this* object *now*'.[7]

To emphasise this contrast between the contents of perception and the contents of thought, I shall henceforward adopt the convention of marking terms referring to perceptual contents with a sub-scripted 'a' for 'analog' – so I shall say that someone has a percept with the content red_a, for example, in relation to which they can apply their recognitional concept with the content *red*.

There is a choice to be made concerning transitive creature-consciousness, failure to notice which may be a potential source of confusion. For we have to decide whether the perceptual state in virtue of which an organism may be said to be transitively-conscious of something must itself be a conscious one (state-conscious – see below). If we say 'Yes' then we shall need to know more about the mouse than merely that it *perceives* the cat if we are to be assured that it is conscious of the cat – we shall need to establish that its percept of the cat is itself a conscious one. If we say 'No', on the other hand, then the mouse's perception of the cat will be sufficient for it to count as conscious of the cat; but we may then have to say that although the mouse is conscious of the cat, the mental state in virtue of which it is so conscious is *not* itself a conscious one!

I think it best to by-pass all danger of confusion here by avoiding the language of transitive creature-consciousness altogether. Nothing of importance would be lost to us by doing this. We can say simply that organism O *observes* or *perceives* X; and we can then assert explicitly, if we wish, that its percept is or is not conscious.

It should be noted that this move is by no means uncontentious, however. For there are some philosophers (notably Dretske, 1995 and Tye, 1995) who think that the notion of transitive creature-consciousness is the basic one, in terms of which the more problematic notion of phenomenal consciousness (see section 3.4 below) is to be explained. Thus

[7] Here is at least part of the source of the common idea that consciousness – in this case, transitive creature-consciousness – is *ineffable*, or involves indescribable properties. But it should be plain that there is nothing especially mysterious or problematic involved. That our percepts have sufficient fineness of grain to slip through the mesh of any conceptual net does not mean that they cannot be wholly accounted for in representational and/or functional terms. I return to this point in chapter 5, where I shall also spend some time discussing the relative primacy of non-conceptual and analog intentional contents in accounting for the nature of our experience.

Dretske, for example, thinks that there is nothing more to the notion of state-consciousness than is already contained in the idea of transitive creature-consciousness – he thinks that it adds nothing to say that a mental state is conscious, beyond saying that the *organism* is, via that state, conscious of something else. These views will be discussed and criticised in chapters 5 and 6.

3.3 Creature-consciousness 3 – self-consciousness

There is one further notion (or rather, as we shall see, *pair* of notions) to be placed on the map – if only to be left to one side – lest it be confused with any of the notions of consciousness already discussed or about to be discussed. This is the notion of *self*-consciousness.

Self-consciousness admits of both weaker and stronger varieties, where each is a dispositional property of the agent. In the weak sense, for a creature to be self-conscious is just for it to be capable of awareness of itself as an *object* distinct from others (and perhaps also capable of awareness of itself *qua* object as having a past and a future). Put differently, the weak form of self-consciousness is a capacity for transitive creature-consciousness, with the self *qua* body as *object* of consciousness.

To be self-conscious in this sense is just to be capable of perceiving and/or thinking of oneself. This weak form of self-consciousness is conceptually not very demanding, and arguably many animals will possess it. Roughly, it just involves knowing the difference between one's own body and the rest of the physical world. And to the extent that transitive creature-consciousness is not particularly challenging or interesting, to that extent self-consciousness, too, can happily be left to one side.[8]

But there is also a stronger notion of self-consciousness, which involves higher-order awareness of oneself *as a self*, as a being with mental states and a subjective inner life. This is much more demanding, and arguably only human beings (together, perhaps, with the other great apes) are self-conscious in this sense. In order for an organism to be self-conscious in this manner, it has to be capable of awareness of itself as an entity with a continuing mental life, with memories of its past experiences, and knowledge of its desires and goals for the future. This is even more demanding than higher-order forms of access-consciousness (see

[8] I do not mean that the notion of bodily self-consciousness is wholly unproblematic, or that there are no questions of interest relating to it. See Bermúdez, 1998, for an interesting discussion of the relationship between indexical self-reference and various forms of non-conceptual and/or non-conscious self-awareness; and see my 1999a for a review. I just mean that the problems, here, do not bear on the issue of phenomenal consciousness, with which this book is primarily concerned.

section 3.5 below), since it involves, not just a capacity for higher-order thought (HOT) about one's current mental states, but a conception of oneself as an on-going entity with such states – that is, with a past and future *mental* life.

The interesting and problematic notion here, for our purposes, will be higher-order access-consciousness (present tensed), which a creature can in principle enjoy *without* having the cognitive sophistication to represent to itself its own past and/or future mental states. To be capable of mental states which are conscious in the higher-order sense, a creature does not need to have a conception of itself as an on-going subject of such states, nor does it need to be capable of attributing past or future states to itself, *qua* self, as subject. It just has to be capable of HOTs about (some of) its states, as and when they occur.[9]

3.4 State-consciousness 1 – phenomenal

The forms of consciousness distinguished and discussed thus far have all of them been properties of the subject of consciousness – it is the person or animal which is conscious *simpliciter*, or conscious of some thing or state X, or self-conscious. The next set of distinctions will now be concerned with forms of consciousness which are properties of mental states. Here it is the mental state of the organism which is said to be conscious or non-conscious, rather than the organism itself.

The most obvious and striking (and the most famous) form of state-consciousness is *phenomenal* consciousness. This is the property which mental states possess when it is *like something* to have them (Nagel's famous phrase, 1974). Put differently, phenomenally conscious states have distinctive subjective *feels*; and some would say they have *qualia* (I shall return to this terminology in a moment).

Most people think that the notion of phenomenal consciousness can only really be explained by example. So we might be asked to reflect on the unique quality of the experience we enjoy when we hear the timbre of a trumpet-blast, or drink-in the pink and orange hues of a sunset, or sniff the sweet heady smell of a rose. In all these cases there is something distinctive which it is *like* to undergo the experience in question; and these are all cases of states which are phenomenally conscious. As Block (1995) puts it: phenomenal consciousness *is* experience.

[9] Again, this isn't to say that there are no questions of interest relating to this demanding form of self-consciousness. On the contrary, there is the question whether self-consciousness (in the strong sense) presupposes awareness of embodiment, in such a way that we can show that any self-conscious creature is essentially embodied, and must have knowledge of its embodiment. See Evans, 1982; Cassam, 1997.

Explanations by example look somewhat less satisfactory, however, once it is allowed that there can be *non*-conscious experiences (see section 3.7 below, briefly, and chapter 6 at length). If there can be experiences which are not conscious ones, then plainly we cannot explain the idea of phenomenal consciousness by identifying it with experience. Perhaps what we *can* say, however, is that phenomenally conscious events are ones for whose properties we can possess introspective recognitional capacities (or at least, ones whose properties are *similar* to those for which we can possess such capacities – the qualification here is introduced to allow for the possible phenomenal consciousness of bats and other organisms with very different perceptual faculties from our own). And then the citing of examples can best be understood as drawing our attention, introspectively, to these properties.

Phenomenally conscious events are ones which we can recognise in ourselves, non-inferentially, or 'straight off', in virtue of the ways in which they feel to us, or the ways in which they present themselves to us subjectively. And note that this need not be construed in such a way as to imply that phenomenally conscious properties depend for their existence upon our recognitional capacities for them – that is, it need not imply any form of higher-order thought (HOT) account of phenomenal consciousness. For it is the *properties recognised* which are phenomenally conscious; and these need not be thought to depend upon our capacities to recognise them.[10]

Note, too, that this talk of what an experience *is like* is not really intended to imply anything relational or comparative. Knowing what a sensation of red *is like* is not supposed to mean knowing that it is like, or resembles, some other experience or property X. Rather, what the experience is *like* is supposed to be an intrinsic property of it – or at least, it is a property which *strikes us as* intrinsic (see chapter 4:1.4), for which we possess an immediate recognitional capacity. Here the point converges with that made in the previous paragraph: the non-metaphorical substance behind the claim that our phenomenally conscious states are ones which are *like something* to possess is that such states possess properties for which we can have recognitional concepts.[11]

[10] So this characterisation of the nature of *feel* does not beg any questions in favour of the sort of dispositionalist HOT theory to be defended in this book. First-order theorists and mysterians can equally say that phenomenally conscious properties (feels) include those properties for which we possess introspective (second-order) recognitional capacities. For they can maintain that, although we do in fact possess recognitional concepts for these properties, the properties in question can exist in the absence of those concepts and are not in any sense created or constituted by them, in the way that (as we shall see in chapter 9:3) dispositionalist HOT theory maintains.

[11] In effect, the terminology of 'subjective *feel*' and 'what-it-is-like' are quasi-technical in nature, having been introduced by philosophers to draw attention to those properties of

It is phenomenal consciousness which is thought to be deeply – perhaps irredeemably – problematic. As we shall see in later chapters, some philosophers hold that the existence of phenomenal consciousness provides a decisive refutation of physicalism, while others think that we shall never be able to understand how phenomenally conscious states *can be* physical (while endorsing something like the general argument mentioned in section 1 above for thinking that they probably are). And many philosophers hold, too, that phenomenal consciousness raises insuperable difficulties for functionalist and theory-theory accounts of the mental. These questions form the subject-matter of the remaining chapters of this book.

An important word about terminology, however, before we proceed. Many philosophers use the term 'qualia' liberally, to refer to those properties of mental states (whatever they may be) in virtue of which the states in question are phenomenally conscious. On this usage 'qualia', 'subjective feel' and 'what-it-is-likeness' are all just notational variants of one another. And on this usage, it is beyond dispute that there are such things as qualia.[12]

I propose, myself, to use the term 'qualia' much more restrictedly (as some other writers use it), to refer to those putative *intrinsic and non-representational* properties of mental states in virtue of which the latter are phenomenally conscious. On this usage, it is not beyond dispute that there are such things as qualia. On the contrary, it will be possible to be a qualia-irrealist (denying that there exist any intrinsic and non-representational properties of phenomenally conscious states) without, of course, denying that there is something which it is *like* to smell a rose, or to undergo a sensation of red or of pain.

3.5 State-consciousness 2 – functional

In addition to phenomenal consciousness, it is possible to distinguish various functionally definable forms of mental-state consciousness. So

our experiences for which we can possess immediate recognitional capacities, or to properties which are relevantly similar to those for which we can possess such capacities (remember the bat).

[12] This is not to say that it is beyond dispute that there exists any such natural property as *the feel of an experience of red*. On the contrary, according to the conception of natural properties to be adopted in chapter 2, it will be an open question whether there are any natural qualia-properties (even in the weak sense of 'qualia'). Rather, qualia-terms might apply – in the manner of terms such as 'spice' and 'sport' – in virtue of the instantiation of a heterogeneous variety of distinct natural properties. What *will* remain beyond dispute is that people sometimes undergo experiences which have the *feel of red*, and that I am undergoing an experience with such a feel right now as I look at my lunch-time tomato. (Compare: it will remain beyond dispute that there are spices, and that paprika is a spice, even though spices do not constitute a natural kind.) In fact, though, on the account to be defended in chapter 9, it will turn out that there *is* a single natural property picked out by terms such as 'feel of an experience of red'.

when we talk about conscious as opposed to non-conscious mental states we *might* have in mind the distinction between states with, and states without, *feel*; but equally, we might have in mind a distinction between states whose occurrence is available to, or known by, the subject, as opposed to states which are *not* so available.

Some use the term *access*-consciousness in this connection (e.g. Block, 1995). But then it is important to distinguish between first-order and higher-order forms of access. A state can be access-conscious in the sense that it is *inferentially promiscuous*, occurring in such a way that its content can figure in the subject's practical and theoretical reasoning and planning, and for expressing in speech. This notion corresponds, very roughly, to what many people think of as *central cognition* (e.g. Fodor, 1983) – as a functional position or mode of occurrence of mental states such that they then can, in principle, interact with any other similarly-occurring states. That is, beliefs can interact with desires to determine intentions, beliefs can interact with other beliefs or with perceptions in generating new inferences, and so on – where all of this activity can be characterised in purely first-order terms.

On the other hand, mental states can be access-conscious in the sense that their *occurrence* is accessible to the subject, in such a way that the subject may be said to know that the states in question exist. A state which is higher-order access-conscious is one that the subject can think *about* as and when it occurs, as opposed to merely helping in the generation of other first-order thoughts. In the human case, of course, these two forms of access-consciousness coincide, at least in fairly large measure. States which are widely available in a first-order way tend also to be available to be thought about by the subject, and vice versa. But still the distinction is an important one to draw for explanatory purposes, as we shall see.

It seems plain that there is nothing deeply problematic about functionally definable notions of mental-state consciousness, from a naturalistic perspective. For mental functions and mental representations are the staple fare of naturalistic accounts of the mind – a point I return to in more detail in chapter 4. But this leaves plenty of room for dispute about whether such notions can help in the explanation of phenomenal consciousness, and about the form which the correct functional account should take. Some claim that for a state to be conscious in the relevant sense is for it to be poised to have an impact on the organism's first-order decision-making processes (Kirk, 1994; Dretske, 1995; Tye, 1995), perhaps also with the additional requirement that those processes should be distinctively *rational* ones (Block, 1995). Others think that the relevant requirement is that the state should be suitably related to higher-order representations of that very state, of various sorts – higher-order thoughts

(HOTs), higher-order linguistic descriptions (HODs), and/or higher-order experiences (HOEs). (See Armstrong, 1984; Dennett, 1991; Rosenthal, 1986; Carruthers, 1996a; Lycan, 1996.)

It is plain that we do *need* some notion of access-consciousness in addition to a notion of phenomenal consciousness, because at least some states can – in a fairly intuitive sense – be conscious without there being anything which it is *like* to undergo them. Consider acts of thinking, in particular. While it may be true as a matter of fact that all conscious acts of thinking have subjective *feel*, because all such acts occur in 'inner speech' or in visual or other forms of imagery (Carruthers, 1996a, 1998c), it does not seem to be part of the very concept of a thought that this should be so. Indeed, as I noted earlier, many people believe that they entertain thoughts which are conscious in the sense that they immediately know themselves to be having them, but where those thoughts are *not* phenomenally conscious.

It appears that there can be states which are access-conscious without being phenomenally conscious. Can there also be states which have subjective *feel* without being accessible to the subject? This is a matter of some dispute, to which we return in chapter 6. In part the answer will turn on whether or not phenomenal consciousness can be *explained in terms of* some notion of access-consciousness and, if so, in terms of *which* notion.

3.6 State-consciousness 3 – standing versus occurrent

An important distinction needs to be drawn between standing (dormant) mental *states*, and occurrent (active) mental *events*. The former category would include beliefs, long-term goals, personal memories, and so on, which one can retain for long periods of time, and even while asleep or comatose. The latter category would include acts of judgement, felt desires, pains, and current perceptions. I propose the following thesis: to say of a standing state – such as a belief, for example – that it is conscious, is to say that it is *apt to emerge in* some appropriate *occurrent* event with the same content which is conscious (in this case an assertoric judgement). So for the belief that grass is green to be conscious, is for me to be apt to *think* (judge) consciously that grass is green when the occasion demands.

It would surely *not* be correct to analyse the conscious status of a standing state directly in terms of some sort of higher-order access-relation to the subject. For it is now familiar that I may be able to know of myself that I have a certain belief or a certain desire without entertaining that belief or desire consciously. That is, I may know by inference from my own behaviour that I believe that *P*, without being disposed to judge, consciously, that *P*. And in that case my standing belief is not, surely, a conscious one.

Nor should we explicate the conscious status of a standing state, such as a belief, by saying that it is one whose existence is *non-inferentially* available to the subject, either. For in fact the way in which we have knowledge of our own dormant beliefs is by first activating those states into an occurrent judgement, and then attributing to ourselves belief in the content of that judgement, as a number of writers have pointed out (Evans, 1982; Gordon, 1995; Peacocke, 1998). So in order to know whether or not I believe that the world is getting warmer, for example, I must first ask myself the first-order question, 'Is the world getting warmer?' If I find myself inclined to answer, 'Yes' (hence activating the first-order judgement, 'The world is getting warmer'), I then embed that content in a report of belief, ascribing to myself the *belief* that the world is getting warmer. The primary thing about a conscious standing state, then, is that it should be apt to emerge in a conscious first-order occurrent event with the same content. Accordingly, then, it is on the conscious status of occurrent mental events that I shall concentrate in what follows.

3.7 Non-conscious mentality

If the notions of state-consciousness so far distinguished are to have any real *bite* or significance, then it must be possible for mental states to be *non*-conscious. It has been a familiar idea at least since the writings of Sigmund Freud – now absorbed and integrated into our folk-psychological conception of the mind – that propositional attitudes such as beliefs and desires can be active in cognition without becoming conscious. But we do not have to buy into the doubtful idea of a *Freudian* unconscious to accept this. The same idea is also accessible by other routes.[13]

Here is one line of thought which makes it seem highly likely that beliefs and desires can be activated without emerging in conscious thought-processes. Consider a chess-player's beliefs about the rules of chess, for example. While playing, those beliefs must surely be activated – organising and helping to explain the moves made, and the pattern of the player's reasoning. But they are not consciously rehearsed. Chess-players

[13] I shall write throughout this book of *non*-conscious as opposed to *un*conscious mental events, precisely to distance myself from any association with the Freudian unconscious, with its commitments to mechanisms of repression, traumas of early childhood sexuality, memory-recovery through analysis, and so on. These ideas are not taken seriously in the cognitive sciences today (although they continue to be influential within the broad area of 'cultural studies'), and the psychotherapeutic practices which they have spawned continue to cause a great deal of harm. Almost the only respect in which Freud's influence has been beneficial, in my view, is in causing ordinary folk to accept the idea of non-conscious mentality, thus leading them closer to the truth, anyway, if no closer to happiness or mental health.

will not consciously think of the rules constraining their play, except when required to explain them to a beginner, or when there is some question about the legality of a move.

The beliefs in question will remain accessible to consciousness, of course – players can, at will, recall and rehearse the rules of the game. So considered as standing states (as dormant beliefs), the beliefs in question are still conscious ones. We have nevertheless shown that beliefs can be non-consciously activated. The same will presumably hold for desires, such as the desire to avoid obstacles which guides my movements while I drive along absent-mindedly (see chapter 6). So thoughts as events, or mental episodes, certainly do not have to be conscious. And then it is by no means redundant to say of a particular such episode that it is a conscious one.

Essentially the same point can be established from a slightly different perspective, by considering the phenomenon of non-conscious problem-solving. Many creative thinkers and writers report that their best ideas appear to come to them 'out of the blue', without conscious reflection (Ghiselin, 1952). Consider, also, some more mundane examples. I might go to bed unable to solve some problem I had been thinking about consciously during the day, and then wake up the next morning with a solution. Or while writing a paper I might be unable to see quite how to construct an argument for the particular conclusion I want, and so might turn my conscious attention to other things. But when I come back to it after an interval, everything then seems to fall smoothly into place. In such cases I must surely have been thinking – deploying and activating the relevant beliefs and desires – but not consciously.

A theoretical case in support of non-conscious thinking can also be made out. For I have already noted above that it doesn't seem to be built into the very idea of a conscious act of thinking, that such an act has a subjective *feel*, or is phenomenally conscious. This is then one of the things which motivates a distinction between phenomenal consciousness and access-consciousness. And it follows, surely, that if the conscious status of an occurrent thought is to be explained in terms of some sort of *access*, then there is no conceptual barrier to the idea that thoughts might be activated *without* being conscious (that is, in the absence of access).

The idea of non-conscious experience, or non-conscious perception, is felt by many to be much more deeply problematic, however. Some people are tempted by the idea that an event can only count as an experience, or as a perception, if it is *like something* to entertain it. So some are inclined to believe that *phenomenality* is intrinsic to the very nature of experience, in which case the phrase 'conscious experience' will be redundant. All perceptual states must be conscious ones, on this view, because all perceptual states must have subjective *feel* or must be *like something* to have.

In chapter 6 I shall argue at some length that such a view is mistaken. What I shall argue is that there are states which are just like conscious percepts in respect of their representational properties and behaviour-guiding causal role, but which are non-conscious, at least in the sense of being inaccessible to their subjects. In which case – if we believe that states which are not access-conscious cannot at the same time remain phenomenally conscious – we should accept that there are perceptual and/or experiential states which are not conscious in either sense. Or alternatively – if we think that states which are not access-conscious can nevertheless have subjective *feel* – we shall have to believe that there are phenomenally conscious perceptual states to which the subjects of those states are blind. I return to these alternatives in chapter 6. For the moment, the question whether phenomenal consciousness implies any contrasting notion of *non*-conscious perceptual states can be left moot.

3.8 Attention and degrees of consciousness

How is the notion of *attention* related to the various notions of consciousness which we have distinguished thus far? There are some who think that the former notion is basic (e.g. Peacocke, 1992). On this account, transitive creature-consciousness of some object or event is really just a matter of the creature *attending* to that object/event. And for a mental state to be conscious is for it to form the content of the creature's attention.

I think, in contrast, that attention is really just an *information-gathering* notion. To say that someone is attending to some object or event is to say that they are directing their sense-organs and/or cognitive resources in such a way as would normally gather rich and detailed perceptual information concerning that object/event. Attention is the process or processes which *select* a given stimulus or input for detailed processing.

The paradigmatic way of attending to something visually is to focus on it, using foveal vision to generate the richest available information concerning it. But cognitive scientists now routinely work with notions of attention which are sub-personal, maintaining that there are a variety of mechanisms unknown to normal subjects which trigger detailed processing, either in a 'bottom-up' way (for example, the loudness of a noise, or the sound of your own name, can grab your attention), or 'top-down', directing the selective processing of information already contained within the perceptual system.[14]

[14] See Kosslyn, 1994, for example, who envisages a kind of expandable 'attentional window' internal to the visual system, which operates top-down to instruct the system to selectively process information from any given region (of whatever size) of the visual field. See also Treisman, 1993, who distinguishes four different forms of visual attention.