

## Consciousness

Consciousness has in recent years become the hot topic among philosophers of mind, and among not a few neuroscientists and cognitive scientists too. The reason has largely to do with the qualia problem surveyed in the last chapter. The received wisdom is that if we distinguish between, on the one hand, the conscious mind's capacity to represent the world beyond itself (that is, its **intentionality**) and **to reason** on the basis of such representations; and on the other, the **qualia** associated with these mental states and processes, then (a) it is the latter - the qualia - rather than rationality or intentionality, that are essential to conscious states qua conscious, and (b) it is these qualia that make consciousness difficult to account for in materialist terms, with rationality and intentionality being readily amenable to a reductionist explanation.

My own suspicion is that this received wisdom has things backwards, on both counts: it is not qualia but the other mental phenomena - rationality and, especially, intentionality - which are essential to consciousness, and which pose the most important challenge to materialism. Ironically, consideration of the views of some contemporary theorists representative of the received wisdom will help us to see this. Their strategy is to give a materialistic explanation of consciousness by first reducing qualitative states (those characterized by qualia) to intentional states (those characterized by intentionality), and then completing their explanation by carrying out (what they suppose to

be the easier task of reducing intentional states to material states of the brain. In this chapter we will examine, among other theories of consciousness, some attempts to develop the first part of this strategy—often called the intentionalist approach - and see that, while none proposed so far is free of difficulties, each of them plausibly contains elements of truth, and can be combined into a general intentionalist account of consciousness. Chapters 5 and 7 will then consider whether intentional mental states and processes really can be accounted for in purely materialistic terms.

## **Eliminativism**

The intentionalist approach to consciousness holds that conscious states are nothing more than intentional states: states exhibiting intentionality or the capacity to represent something beyond themselves. The difficulty with this approach is that qualia seem devoid of intentionality: the throb of a toothache, for example, doesn't seem to represent anything; it just hurts. So qualia seem to be an extra element, an aspect of conscious experiences over and above their intentional content. The overall experience of a toothache may include—the thought that one is in pain - a thought which, representing as it does one's current situation, exhibits intentionality - but the pain itself is a further, non-intentional, component. Conscious experiences, therefore, cannot be completely reduced to intentional states. In particular, qualia are irreducible to intentional properties, and must somehow be accounted for separately, independently of any materialist analysis of intentionality.

Daniel Dennett's response to this difficulty is, whatever else one might say about it, bold: he simply denies that there really are any qualia to account for in the first place. His is what philosophers call an eliminativist position, one that deals with a philosophically problematic phenomenon by suggesting that its problematic nature gives us reason to doubt its existence - to "eliminate" it entirely from our picture of the world, rather than attempting to explain it. He does not deny that we really do have conscious experiences - feeling pain, tasting coffee,

smelling flowers, hearing music, and all the rest – but denies only that any of these experiences feature properties of the sort qualia are taken to be. There are, that is to say, no properties that are essentially intrinsic - that is, unanalyzable in terms of their relations; or subjective - that is, directly accessible only from the first-person point of view. The throb of a toothache, appearances notwithstanding,

### *Consciousness 93*

is neither of these things. It was suggested in the previous chapter that qualia might not be essentially intrinsic in the sense they are often claimed to be; to this extent Dennett may be right. But it was also suggested that they do seem to be essentially subjective. So what of Dennett's claim that there are no essentially subjective properties? Isn't it just obviously false, given what we know from introspection?

Recall from chapter 3 that materialists often take our commonsense concept of the mind to constitute a kind of theory that can be described as "folk psychology." If one grants this assumption, then the entities supposedly "postulated" by folk psychology- such as qualia - count as theoretical entities: they might turn out to exist, as the best explanation of the phenomena they are postulated to explain; but then again, they might turn out not to exist, for there might be a better explanation that does not postulate them. But even if we do grant this, is there really any reason to doubt that qualia, even if theoretical, are real? Dennett thinks there is, and in defending his eliminativism he revisits the sort of qualia inversion scenarios considered in the last chapter. Suppose you wake up after neurosurgery and are baffled to find that grass looks red, and the sky looks yellow. It might seem obvious that your color qualia had been inverted, presumably due to some playful rewiring of your neurons. But, as Dennett argues, that is not the only possibility. The neurosurgeons might have produced your bafflement by tampering with whatever neural connections underlie your perceptions of color, thereby inverting your qualia, but they might instead have done it by tampering with the connections underlying memory: maybe your qualia are the same now as they always have been, and you are only misremembering how they seemed before. The only way you could possibly determine which of these possibilities is actual is by asking the neurosurgeons or,

perhaps, doing some sort of neurological selfinspection. But then you must necessarily rely on objective, third-person evidence to know whether your qualia have been inverted; and in that case, Dennett says, qualia can't be subjective. But if qualia are held to be essentially subjective - subjectivity being part of their very essence then this just entails that there really are no qualia. Whatever the inverted spectrum scenario, and color vision in general, involve, they do not involve the having of qualia, and we ought therefore to prefer a theory of mind that does not make reference to qualia.

One could object that this argument appears to be a non sequitur. That whether your memory of your qualia has been tampered with is something you need to appeal to third-person neurological evidence to determine does not seem to show that your qualia themselves - Past or

#### 94 *Philosophy of Mind*

present - can be known only by appealing to that evidence. You might, for all Dennett has said, still be directly aware of your qualia from the first-person, subjective point of view even if you don't know whether they are the same as or different from the sort of qualia you had yesterday - just as you might really be aware of the book in front of you even if you don't know whether it was the same as or different from the book you saw yesterday. Questions about memory do not necessarily have a bearing on the nature of your awareness of objects present here and now (even if they have an obvious bearing on what you can justifiably claim to know about such objects), whatever those objects happen to be.

Of course, the analogy isn't exact. There is no doubt that you really are aware of your qualia now even if you don't know whether or not they're like the ones you had yesterday; in the case of the book, you might not really be aware of it right now, for you might be merely hallucinating it. And if the indirect realist theory discussed in chapter 1 is correct, then even if you are aware of it, you are not aware of it directly, in the way you are aware of your qualia. But all this seems only to strengthen the suggested **reply to Dennett**. For, if indirect realism is correct, it is only through the first-person, subjective realm of qualia that we know that there is an objective, third-person realm – including neurosurgeons and the brains they might tamper with - in the first place.

Indeed, puzzles concerning memory of the sort Dennett makes use of, when one pushes through their implications consistently, serve to underline (rather than undermine) the reality of the first-person, subjective realm of qualia: that the entire Past is a figment of my imagination, and the universe really only five minutes old, is yet another skeptical scenario of the sort considered in chapter 1, one raised this time by consideration of the possibility of faulty memory. Nor will appeal to third-person neurological evidence by itself serve to refute such skeptical worries, for such an appeal would itself assume the reliability of one's memory (that is, it would assume that one was correctly remembering what the neurologists had told one or what one had read in textbooks about the links between certain neural structures and memory). So even to trust the evidence from the neurosurgery requires first being able to show you can trust the subjective evidence, of your senses, via arguments (of the sort also considered in chapter 1) that can themselves be defended entirely from the first-person point of view.

It seems we ought, for these reasons, also to reject the assumption that qualia are theoretical entities in the first place. Far from being the

### *Consciousness 95*

postulates of a theory they are, rather, among the data to which all empirical theorizing and postulating must appeal. Dennett would object that appeal to such first-person, subjective data is incompatible with the objectivity demanded by scientific method. He holds, accordingly, that only evidence available from the third-person objective point of view ought to form the basis of a scientifically respectable theory of the mind. Given such a constraint, materialism, and indeed eliminativism, seem to follow automatically, even trivially. But to insist on this constraint seems, by the same token, simply to beg all the important questions. It is also to take a position that is prima facie implausible, especially if one accepts the indirect realist view considered in chapter 1. In any case, Dennett's assertion that scientific objectivity requires appealing exclusively to third-person evidence appears mistaken. It certainly would have come as a surprise to a thinker like Carnap, whose regard for science as the touchstone of objective knowledge was legendary (indeed, legendarily excessive), yet who regarded respect for the first-person (or, as he called

it, autopsychological) point of view as fully consistent with such objectivity. What scientific objectivity requires is, not denial of the first-person subjective point of view, but rather a means of communicating intersubjectively about what one can grasp only from that point of view. Given the relational structure first-person phenomena like qualia appear to exhibit - a structure that, as we saw in the last chapter, Carnap devoted great effort to elucidating - such a means seems available: we can communicate what we know about qualia in terms of their structural relations to one another. Dennett's position rests on a failure to see that qualia being essentially subjective is fully compatible with their being relational or non-intrinsic, and thus communicable. This communicability ensures that claims about qualia are epistemologically objective, that is, they can in principle be grasped and evaluated by all competent observers, even though they are claims about phenomena that are arguably not metaphysically objective, that is, they are about entities that exist only as grasped by a subject of experience. It is only the former sort of objectivity that science requires. It does not require the latter - and cannot plausibly require it if the first-person realm of qualia is what we know better than anything else.

## **Representationalism and Higher-order Theories**

If qualia cannot be dismissed as unreal, then, how can an intentionalist theory of consciousness deal with them? The most straightforward

### *96 Philosophy of Mind*

answer is representationalism, the view that qualia are nothing more than representational properties of conscious experiences. The redness of your experience of seeing an apple, for instance, is just a representation of the objective redness of the apple itself, of the physical property of the surface of the skin of the apple by virtue of which it absorbs some wavelengths of light and reflects others. There is, on this view, nothing more to the redness than that its intentionality or representational content is all the content it has, and there is no distinctly qualitative element over and above that. So, the problem of qualia reduces to the problem of intentionality; it does not pose a separate challenge to materialism.

What about bodily sensations that do not seem to have such representational content? To return to the example of a **toothache**, its nagging quality does not seem to represent anything; it appears to be nothing more than what philosophers sometimes call a "**raw feel**," a pure sensation without any intentionality or meaning (even though, again, one's thoughts about the pain would of course have intentionality or meaning). But the representationalist would hold that such cases are not genuine counter-examples. The qualia associated with toothache can plausibly be taken to represent something, namely the damage to the tooth that causes the toothache. By the same token, pains in general can be taken to represent damage to the parts of the body in which they are felt, and other bodily sensations can be taken to represent other states of the body.

Even if we accept all this, there is still the problem of accounting for why representational states like seeing an apple or feeling pain are associated with consciousness, while other representational states (for example your belief that  $2 + 2 = 4$  which you have even when you are not conscious of it) are unconscious. If to be a conscious experience is just to be a state having a certain representational content, wouldn't all states with representational content be conscious? But they aren't all conscious; so some extra element, in addition to their representational content, must be what makes certain states with representational content conscious, and representationalism thus cannot be the full story about consciousness.

Here is where some philosophers would appeal to a **higher-order theory of consciousness**. The idea here is that what makes any particular mental state a conscious state is that it is the object of a higher-order mental state that represents it. Some versions of this theory would take such higher-order states to be **thoughts**, while others would take them to be more akin to **perceptions**: in the first version,

*Consciousness 97*

just as one might have a thought about some object in the external world, one might also have a thought about a thought, or about some other kind of mental state; in the second, just as one might have a perception of an object in the external world, one might also have an "inner" perception of the perception itself.

The overall picture of consciousness that emerges from these theories is this: what gives a particular conscious experience the particular qualitative character it has - that is, what makes it the case that it is associated with particular qualia - is the unique representational content embodied in those qualia. Some theorists would also add that the structural relations, alluded to above and discussed in the previous chapter, by which each quale can be uniquely identified in terms of its similarities and dissimilarities to other qualia, also play a role in determining the precise character of a conscious experience. But representational content and/or structural relations between qualia, even if they can account for why an experience has this qualitative character rather than that, still do not explain why it has any such character at all. To explain that requires appeal to a higher-order account: a state is conscious when there is another, higher-order state which represents it. The Presence of such a higher-order state thus ensures that the particular mental state represented by it counts as a conscious experience; and the elements of that conscious experience having the particular representational content and/or structural relations they do ensures that it is a conscious experience of this sort rather than that.

There is much to be said for this approach (or combination of approaches), but it seems insufficient as it stands. Representationalists and higher-order theorists (and structural relation theorists like Clark and Hardin too, for that matter) generally see their accounts as variations on functionalism: representational states and higher-order states are interpreted by them as fully analyzable in terms of the causal relations they bear to stimulation of the sensory organs, other internal states, and behavior. But then their accounts would appear to be as vulnerable to the anti-materialist arguments of the previous chapter as is any other version of functionalism. For example, a **zombie** duplicate of you would not only have an internal state caused by light reflected from an apple striking its retinas, signals from the retinas being sent to the visual centers of the brain, and so on, but would also have a further ("higher-order") internal state caused by the first internal state, and all these states together would produce behaviors like salivating, or say "Look, an apple!"; yet such a zombie would, nevertheless, lack any



subjective conscious experience of the apple. So, the notion of higher-order mental states, understood in functionalist terms, appears to add little to a materialist account of consciousness.

If representationalist and higher-order theories are to shed new light on the problems of consciousness and qualia, then, it seems they must somehow go beyond the standard functionalism in which they are usually embedded. To see one way in which this might be accomplished requires a digression

### **Russellian identity theory and neutral monism**

Thus far in this book we have focused on dualism and materialism as the main alternative general metaphysical approaches in the philosophy of mind. That is, we have considered the views that everything is ultimately material (materialism), and that the material and the mental are equally ultimate (dualism). These alternatives are paid the most attention by contemporary philosophers of mind, but they are not the only alternatives to be proposed in the history of the subject. A third view, known as idealism, holds that everything is ultimately mental - for example, the version associated with George Berkeley (1685-1753) holds that purportedly physical objects like tables and chairs really exist only in so far as a mind perceives them to exist. But though idealism has had some illustrious defenders in the history of philosophy, it is not generally regarded as a serious option by most contemporary philosophers (with some important exceptions). There are two other, more promising, alternatives that we will be exploring, one in this chapter and the other in chapter 8. The first holds that neither mind nor matter is metaphysically ultimate: what is ultimate is rather a single kind of stuff that is neutral between, and more fundamental than, either of them. This is, in a nutshell, the metaphysical theory known as **neutral monism**.

The most important proponent of this view in the twentieth century was **Bertrand Russell**. His formulation of it evolved significantly through the course of his long career; what we want to focus on is the final, settled version. Russell begins by drawing out the implications of the **indirect realism** he endorsed, and which we discussed in chapter 1. If

in perception we are directly aware, not of external physical objects themselves, but rather only representations of those objects, then we have in Russell's view no grounds for supposing that those objects really have the properties they are presented to us by perception as having. We have no reason to assume, for example, that

*Consciousness 99*

the redness and sweetness of the apples we perceive is really in the apples themselves, as opposed to being merely an artefact of our perceptual machinery - just as the redness you see on the wall in front of you when you are wearing glasses with red lenses is, for all you know, not really in the wall itself but only an artefact of the glasses. As we've noted before, physics seems to give us positive reason to believe that the redness and sweetness are not in the apples: for like every other physical object, an apple is in reality nothing but a collection of colorless, odorless, tasteless particles. What the physical world is really like "in itself," apart from our perceptual representations of it, is not something perception can tell us.

What does tell us what the physical world is really like is science. But science, Russell argues, does not tell us nearly as much as we often assume it does. For instance, what exactly are these colorless, odorless, tasteless particles of which physics speaks - molecules, atoms, quarks, gluons and so forth? Physics defines these entities entirely in terms of their causal relations to one another: a molecule is whatever plays such-and-such a causal role at the microscopic level, an atom is, among other things, what plays the role of serving as a component of a molecule, and so on. But what exactly it is that happens to play these roles is something physics does not tell us. We know from science only that the material world is a collection of fundamental entities having a certain causal structure, a structure described in mathematically precise detail by the physical sciences; but what it is that fleshes out this causal structure, the intrinsic nature of the specific entities that bear these causal relations to one another by filling out each place in the vast causal network described by science, is something we do not know. (This is a view about the nature of scientific knowledge known as **structural realism**: realist because it holds that there really is a physical world existing external to our minds,

structuralist because it holds that all we know of that world is its structure rather than **intrinsic nature**.)

Our knowledge of the external physical world turns out to be highly abstract; including our knowledge of the brain, considered as the object of neuroscientific research, as one external physical thing among others. The brain is not in reality the greyish, squishy thing we encounter in perception: that is only a subjective, perceptual representation of the brain. The brain is, rather, a complex causal structure of neural events, where these neural events are defined in terms of their characteristic causes and effects rather than in terms of the qualities presented to us in visual or tactile inspection of the brain. The **inner nature** of what specifically has these cause and effect relations is something we do not know –

### 100 *Philosophy of Mind*

or at least, we do not know from either perception or neuroscientific study.

But are perception and scientific inquiry (whether neuroscience, physics, chemistry or whatever) the only possible sources of knowledge about the nature of the brain? Russell suggests that there is one further possibility: **introspection**. In introspecting or looking within itself, the mind is directly aware of its own contents - of thoughts, experiences, and their associated qualia. As materialists have argued, there are, at least in general, correlations between various mental events on the one hand and brain events on the other. Perhaps in introspecting these mental events, and in particular our qualia, we are directly aware of precisely the inner natures of the entities that play the causal roles specified by neuroscience. Perhaps neural events just are the thoughts, qualia, and so forth encountered in introspection. In being immediately aware of the taste of an apple or a sensation of pain, maybe what we're directly aware of are events occurring in the brain, as it really is "in itself."

This is obviously a **mind-brain identity theory**. But it is not the materialist kind of identity theory discussed in chapter 3. Materialism in general seems to take it for granted that we know exactly what the intrinsic nature of the physical world is, and seems to assume also - especially in the case of functionalism - that we do not know (or at least

that pre-philosophical and pre-scientific common sense does not know) what is the intrinsic nature of the mental realm: the function- alist claims that mental states and processes are to be defined entirely in terms of their causes and effects. Russell's view is that this has things precisely backwards. It is in fact the mental world that we know most directly and intimately, and the external physical world that we grasp only in terms of its causal structure. In identifying the mind and the brain, Russell is not, as the materialist identity theorist is, reducing the mind to the brain; if anything it is the other way around. The brain turns out to be the mind; more exactly, the neural events and processes defined only abstractly, in causal terms, by neuroscience turn out to be nothing other than mental events and processes - thoughts, experiences, and the like. The grey squishy thing you've seen pictures of in textbooks or that a neurologist looks at when doing surgery is not what the brain is really like intrinsically. If you want to know what it is really like, you need only focus on the qualia you're experiencing right now. The whiteness and blackness of the paper and ink of the book you're reading, the colors on the cover, the smell and warmth of the coffee in the cup beside you, the feel of your back against the chair: those are the

*Consciousness 101*

brain's true qualities. In introspecting these qualia, you are directly aware of nothing other than the inner nature of your own brain.or, as , Russell paradoxically put it: "I should say that what the physiologist sees when he looks at a brain is part of his own brain, not part of-the brain he is examining"!

If this sounds strange, it is supposed to. But it makes perfect sense when one combines indirect realism with the mind-brain identity thesis. For what Russell means is that the physiologist is not directly aware of the (patient's) brain he is examining, though of course he is aware of it indirectly; what he is directly aware of is a constellation of qualia - greyishness, squishiness, etc. - which are, given the identity theory, identical to features of his own brain, and which are ultimately a distant effect of the light reflected from the patient's brain traveling to the physiologist's retinas, which sets up a sequence of neural firing patterns eventually culminating in the visual experience. still, the theory definitely

counts as a revision of common sense. More importantly, for our purposes, it counts as a rejection of materialism, for, both epistemologically and metaphysically, it gives priority to the subjective, first-person realm of qualia rather than the objective thirdperson external physical world. Yet it also seems to count as a rejection of dualism, in so far as it identifies the brain with the mind, rather than seeing them as distinct substances.

Indeed, it might seem at first glance to lead instead to a kind of idealism: for if qualia are the intrinsic qualities of the brain, and the brain is - as far as we know from science - made of exactly the same kind of stuff as everything else in the physical universe, wouldn't this entail that everything else in that universe also has qualia as intrinsic qualities? wouldn't qualia be what ultimately make up tables, chairs, rocks, trees, and every other object of everyday experience? If so, this would seem to entail that, in some sense, everything physical is really mental, which is precisely what idealism claims. But Russell and some other philosophers who have endorsed and developed his position, such as Michael Lockwood, have resisted this conclusion. They have suggested that what contemporary philosophers have come to call qualia (this was not Russell's own expression) - reddishness, the nagging character of pain, the pungency of an odor - may well indeed be the intrinsic properties of every physical thing; but they have also suggested that these properties are, contrary to the standard view, not in fact essentially mental properties at all. Reddishness and all the rest need not necessarily exist in the mind of an experiencing subject: they can exist unsensed by any mind, and do so exist when they enter into

## 102 *Philosophy of Mind*

the constitution of physical objects other than the brain. The Russellian view is thus interpreted - at least by Russell himself and Russellians like Lockwood - as a version of **neutral monism**: qualia comprise the single ultimate kind of stuff out of which everything in the world is composed (hence "monism"), but they are intrinsically neither mental nor non-mental (hence "neutral"); they count as mental only when organized into the sort of causal structure described by neuroscience (that is, a brain), and count as non-mental when Organized into other sorts of causal

structures (rocks, trees, tables, chairs, galaxies). Since it identifies qualia with properties of the brain, this account is also a kind of identity theory - sometimes labeled the Russellian identity theory, to distinguish it from materialist identity theories of the sort described in chapter 3.

One of the advantages of this theory, whatever one wishes to call it, is that it seems to be immune to the sorts of objections that, as we've seen, plague materialist theories. In response to the zombie argument, for instance, the Russellian can hold that zombies can be shown not truly to be conceivable when one's exercise in conception is informed by indirect realism (and the structural realism Russell conjoins to indirect realism). Zombies seem conceivable only if, when imagining them to be "physically identical to us," we imagine their brains being the greyish, squishy things we encounter in perception. But of course, to imagine that sort of thing is really only to imagine a perceptual representation of a brain; it no more involves imagining the brain as it really is intrinsically than does imagining a linguistic representation like the word "brain." To note that a greyish, squishy thing can be imagined to exist apart from qualia no more undermines a mind-brain identity theory than the fact that you can imagine the symbol "H<sub>2</sub>O" existing in the absence of water undermines the claim that water = H<sub>2</sub>O. Really to imagine the brain as it is "in itself" would, on the Russellian view, require imagining it as constituted by qualia. But to imagine that is, by definition, not to imagine a zombie, since a zombie is supposed to be a creature devoid of qualia. In that case, however, zombies turn out to be inconceivable after all.

## **Troubles with Russellianism**

Or do they? A number of philosophers take the Russellian position - long neglected in the philosophy of mind, but in recent years making something of a comeback - to be a great advance over the standard alternatives. But arguably, it will not do as it stands. First, the

*Consciousness* 103

suggestion that qualia can exist independently of any experiencing conscious subject is highly counter-intuitive, indeed highly implausible. The very notion of qualia is, after all, introduced as the notion of

properties of immediate conscious experience. So it is questionable whether we can coherently abstract away from the notion of qualia the presence of a conscious subject, a mind, to whom they are presented.

Some philosophers sympathetic with the Russellian approach, such as David Chalmers, acknowledge that qualia require a conscious subject for their existence - and thereby accept the idealism (or panpsychism, as they often prefer to call it, to distinguish their view from the sort of idealism associated with Berkeley) to which this commits them. They don't hold that qualia quite like ours - pains, itches, color sensations, odors, and the like - make up the physical universe outside our minds, for our qualia are no doubt more complex, given the complexity of our brains. At the level of molecules, atoms, and subatomic particles, there are instead what might be called proto-qualia playing the relevant causal roles, properties simpler than, and only vaguely analogous to, our qualia. Associated with these proto qualia, and thus with molecules, atoms, and subatomic particles, would have to be proto-subjects - simple, tiny minds (or proto-minds) having extremely simple experiences (or proto-experiences). It is only when these proto-qualia get organized into highly complex structures like our nervous systems that they somehow, in combination, give rise to complex minds like our own.

The initial, uncharitable objection to all of this is that it is just plain crazy, and Chalmers' critics have not been shy about raising it. For most philosophers, if a theory has implications as bizarre as that basic physical particles are associated with minds (proto- or otherwise) experiencing qualia (proto- or otherwise), that is reason enough to reject it. A more technical objection is that it is hard to see how proto qualia could combine in such a manner as to "add up to" the sort of conscious experience we're familiar with in everyday life - an experience which seems to be a single conscious experience rather than a composite of billions of tiny proto-experiences, and which is present to a single conscious subject rather than to a collection of billions of tiny proto-subjects. A conscious experience, that is to say, has a unified character it would not have if it were an aggregate of simpler elements.

We will return later to the question of the unity of consciousness - a question which by no means poses a challenge to panpsychism alone. Its potentially panpsychist implications are, in any case, not the only problem for the Russellian theory. For it seems that the theory does not

in fact avoid the zombie argument the way some of its defenders seem to think it does. Recall that what is essential to a molecule, atom, or subatomic particle qua molecule, atom, or subatomic particle is, in the Russellian view, that it plays a certain causal role, the role assigned to it in theoretical physics. The Russellian believes that qualia or proto qualia are what play these roles. But could something else have played them instead? There seems no reason not to think so. An analogy might help: what is essential to the particular philosophy professor Feser qua being a philosophy professor is that he is capable of teaching certain classes, directing students in their research, etc. Could someone other than Feser have performed those functions just as well? Much as he'd like to think otherwise, it is true that someone could. There is nothing about Feser qua Feser that is necessary to playing the role of being a philosophy professor: plenty of non-Fesers can and do play the role just as well. Similarly, there seems to be nothing about a quale or proto-quale qua proto-quale that is necessary to performing the functions of a basic physical particle. Something other than a proto-quale, something absolutely devoid of anything even vaguely analogous to qualitative character, could play the role just as well.

This would seem to entail that it really is perfectly possible for there to be a creature physical-particle-for-physical-particle identical to you which is utterly devoid of proto-qualia, and thus of qualia - a creature which has something other than proto-qualia playing the relevant causal roles. But then such a creature would be a zombie, in which case zombies really are conceivable even on the Russellian view. And if that is so, then even the Russellian view entails a kind of dualism: for it entails that qualia are one kind of thing, and the basic physical components of the universe qua physical (that is, quahaving the causal properties described by physical science), which can exist either with or without qualia, are another. Indeed, though Russell and Lockwood take themselves to be identity theorists of a sort, Chalmers does not, and explicitly presents his own panpsychist brand of Russellianism as a version of property dualism.

Would a Russellian property dualism, like other forms of property dualism, be threatened with epiphenomenalism? At first glance, it might



seem not: if qualia or proto-qualia are what play the causal roles physics associates with molecules, atoms, subatomic particles, etc., then they might indeed appear just obviously to have a causal influence on the physical world. But appearances are deceiving. Given that something other than proto-qualia could equally well play those same roles, there is nothing about their distinctly mental, qualitative character that

*Consciousness 105*

is relevant to their playing it. Feser is a husband and father, but his being a husband and father is completely irrelevant to his playing the, role of a professor: someone who was neither a husband nor a father could play that role in exactly the same way. So Feser's being a husband and father is, we might say, epiphenomenal relative to his effects on the world qua philosophy professor. Similarly, a proto-qualia's qualitative character - being proto-reddish, or proto-pungent - is completely irrelevant to its playing the role of a subatomic particle: something lacking proto-reddishness or proto-pungency could have played the role in exactly the same way, so that these proto-qualitative features are epiphenomenal. So not only does the Russellian view lead to property dualism, but it seems to lead to epiphenomenalism too - with all the problems we've seen that entails.

### **A more consistent Russellianism**

Despite these problems Russell's theory might yet prove to be an advance over the usual alternatives. The reason lies not in the theory's metaphysical component - taking qualia to be the intrinsic properties of the material world, with all the weirdness this seems to lead to - but rather in its epistemology, its account of the nature of perceptual knowledge. Russell's central insight was, arguably, to see that indirect realism has dramatic implications for the mind-body problem; but it may have been an insight neither he nor his followers have taken seriously enough, or far enough.

Russell's own defense of indirect realism emphasized the causal element in perception, the way in which all our experiences of the external world are mediated by causal chains. The gap represented by

these chains - by, for instance, the myriad neural firing patterns, retinal cell activity and stream of photons that come between the surface of an apple and your experience of it - entails, in his view, that you never directly get at external objects themselves, but at best only at mental representations of them. Russell assumed, however, that you do indeed, in introspection, directly get at these representations themselves. But do you?

In Russell's view, those perceptual representations are, like all other mental states, identical with certain brain processes, which come at the end of a long causal chain beginning with the surface of an external object. But then the introspection of these representations must be as dependent on the causal workings of the brain as perception is. If your perception of external objects is mediated by causal chains, surely so is

### 106 *Philosophy of Mind*

your introspection of those perceptions, as brain events subserving perception, triggered by impulses from the sensory organs, in turn trigger further brain events subserving introspection. As with perception, introspection would thus seem to provide you with only a representation - an introspective representation - of what you are made aware of through it. It gives you a representation, that is to say, of your perceptual representations themselves; it does not acquaint you with the intrinsic nature of those representations. And if we imagine yet higher-order mental events directed on to introspection itself instances of meta-introspection, if you will - then these too must, on the Russellian model, be regarded as involving yet further causal chains and thus yet higher-level representations (that is, representations of representations of representations).

If this is right; then there is reason to believe that we have, contrary to Russell, no more knowledge of the inner world of the brain as it is "in itself" than we have knowledge of the external physical world as it is in itself. All such knowledge would be mediated by representations. One consequence of this seems to be that the Russellian response to the zombie argument can be salvaged after all. Zombies really are inconceivable, for in conceiving of perceptual experiences and qualia as I encounter them in introspection existing apart from the abstract causal

structure of the brain (or whatever), I am not conceiving of those experiences and qualia as they are in themselves, but only of introspective representations of them. As with Russell's original proposal, we can conclude that conceiving of that sort of thing existing apart from the brain is of no more consequence than is the fact that the symbol "H<sub>2</sub>O" can be imagined to exist in the absence of water. This would also appear to restore to the Russellian view its status as a version of neutral monism rather than property dualism. There is, at least where the question of the relationship between consciousness and the brain is concerned, only one kind of stuff, but it is intrinsically neither mental nor material. We count it as material when it is presented to us via perception, and as mental when presented to us via introspection: hence the brain seems "material" when one examines it during brain surgery, but "mental" when one "looks within" at thoughts, experiences, and feelings; but one is aware of exactly the same object in both cases. The difference between material processes and qualia is a difference only in how we represent things, not a difference in the things themselves as they exist independently of us. It is, that is to say, an epistemological difference, not a metaphysical one.

*Consciousness 107*

### **Consciousness, Intentionality, and subjectivity**

When the Russellian view is modified in the way suggested, we have a position that is in many respects reminiscent of the representationalist and higher-order theories considered earlier: the features we are introspectively aware of as qualia are just features of perceptual representational states, and features of those states, not intrinsically, but only as represented by yet higher-order representational states. Unlike other versions of those theories, this one is not a materialistic functionalist account, since it does not try to reduce qualia to features of objective, third-person material phenomena, and it is therefore not subject to the usual objections to functionalism and materialism.

Of course, this still leaves us needing to explain representation or intentionality itself. But if the problem of qualia can indeed be reduced to the problem of intentionality, that is no mean achievement. And the

other common objections to the intentionalist account do seem answerable. The question of how intentionalism can deal with intentional states that are not conscious - such as one's belief that  $2 + 2 = 4$ , of which one is usually not conscious - is best dealt with by denying the assumption that there are such states in the first place. As John Searle has argued, strictly speaking there really are no processes that are both totally unconscious and literally intentional; rather, what exist are nonintentional, unconscious processes - neural wiring patterns, say - which have come into existence as a result of past learning (for example, one's study of basic arithmetic) and which have a tendency under the right circumstances (for example, when one is balancing one's checkbook) to cause certain states which are both intentional and conscious, such as the conscious belief that  $2 + 2 = 4$ . Searle's reasons for endorsing this connection principle (the connection in question being an inherent connection between intentionality and consciousness) can only be fully understood after we have more closely examined the issues surrounding intentionality; but the principle shows that the objection from so-called unconscious intentional states is hardly fatal.

Intentionalism is also plausible for reasons other than those already considered. As Tim Crane has argued, the essential features of an intentional state include directedness on an object, and what he calls (following Searle) aspectual shape, or the object's being presented in a certain aspect or in a certain way: thinking about the 43rd President of the United States involves your mind's being directed upon a particular man and considering him as the President (rather than as the former

Governor of Texas or the son of a previous president). But conscious states characterized by qualia seem to involve exactly these features. To have a toothache, for instance, is for your mind to be directed upon a particular part of the body - your tooth - and in a certain aspect - as hurting. Furthermore, in both intentional states and conscious states, subjectivity is essential. The directedness of an intentional mental state is always the directedness of the mind of a subject upon an object of thought, and aspectual shape is always the way that object is presented to that subject; similarly, qualitative conscious states always involve things appearing or seeming a certain way to a subject, where the qualia determining the character of that appearing or seeming (such as the particular shape of the reddish patch of color you see when you look at a tomato) always reflect the perspective or point of view of a particular subject (who is, say, to the left of the tomato).

The centrality of intentionality to consciousness and of subjectivity to both is made more evident by a consideration of the unity of consciousness. Consider the experience you're having right now: you see and feel a book and your hands holding it, perhaps against the background of a table, and hear the rustling of the pages as you turn them. We know from modern neuroscience that discrete processes in the brain register each aspect of the physical world you are experiencing - the colors, shapes, and sounds, the motion of the book's pages, the feel of their texture, and so forth, are each correlated with a different neural event. Yet the experience you are having is neither an incoherent jumble of distinct and disconnected features (pages, ink, motion, colors, etc.) nor is it a collection of distinct and disconnected experiences of distinct and disconnected features; it is a single, unified experience of a book, the hands holding it, and a table. The experience has a coherent significance or meaning, and significance or meaning for a single subject of experience. You are not only aware of the shape, texture, colors, etc. as separate elements, but are aware of them as a book; and it is you who are aware of them, rather than myriad neural events somehow each being "aware" of one particular aspect of the book. In this unity of conscious experience, we see again how deeply tied consciousness is to

intentionality, and how both consciousness and intentionality are tied to the presence of a subject.

The overall view suggested by the considerations adduced in this and the previous chapter is this. In perceptual experiences, the conscious subject represents the world external to the mind, and in introspection of those perceptual experiences, the subject represents those experiences themselves. In the first case, the subject is only indirectly

### *Consciousness 109*

aware of the external world; in the second, he or she is only indirectly aware of the perceptual experiences. In both cases, the subject is directly aware of a representation: in the former a first-order representation (of the external world), in the latter a second-order representation (of the first-order representation). In the latter, the first-order representation is represented as being, in various ways, more or less similar to other representations - that is, it is represented as exhibiting certain qualia, where qualia are analyzed in terms of their similarity relations to each other. In so far as conscious experiences, whether first-order perceptual ones or higher-order introspective ones, are ultimately representational, consciousness is at bottom a manifestation of intentionality; in so far as intentionality in general and qualitative similarity judgments in particular require the presence of a subject, and in so far as the indirectness of perception and introspection entail the primacy of the first-person point of view, consciousness-cumintentionality appears to be inherently and irreducibly subjective.

Despite the advances in our understanding of consciousness made possible by the theories examined in this chapter we seem left, metaphysically, in much the same position we found ourselves at the end of the previous chapter: with subjectivity laying at the core of the mental, and persisting as the main obstacle in the way of a materialist account of conscious experience. There is, as we've seen, a sense in which qualitative conscious states might be identified with states of the brain: perception of a brain state and introspection of a mental state can be seen as two different ways of representing the same thing. Still, since the characteristically "material" and "mental" aspects of this thing, whatever it is, turn out to exist not in the thing itself but only in the subject's

representations of it, the sense in which the mental and physical can be identified would be a neutral monist sense, not a materialist sense. Moreover, the metaphysical status of the subject who does the representing of these conscious states/brain states has yet to be determined; in particular, nothing said in this chapter adds plausibility to the suggestion that this representing subject is material in nature.

## **The binding problem**

These matters have not been settled conclusively in favor of the dualist. For, if it is true that the problem of consciousness cannot be divorced from the problem of intentionality, the question of whether materialism can account for subjectivity cannot ultimately be answered until we consider whether it can account for intentionality.

### *110 PhilosoPhy of Mind*

providing such an account will be difficult, as evidenced by what was said earlier about the unity of consciousness. We noted that though the various aspect of the scene you experience are separately encoded by distinct processes in the brain, your experience is, nevertheless, unified: it is an experience of the book, hands, and table all together, and of the book, hands, and table as book, hands, and table rather than as a meaningless sequence of colors, shapes, textures, and sounds. But how exactly is this possible? How do discrete brain processes manage to add up to a meaningful, unified experience?

This is known among neuroscientists, cognitive scientists, and philosophers of mind as the binding problem; while it is often discussed as if it reflected merely a temporary gap in our scientific knowledge, william Hasker has argued (following leads found in the writings of Descartes, Leibniz, and Kant) that it is most likely impossible in principle to be a materialistic, neuroscientific, solution to it. Even if each of the processes in the brain encoding different aspects of the experienced objects were somehow individually conscious (in a manner reminiscent of Chalmers' panpsychism) this brain process conscious of this shape, that process conscious of that color, a further process conscious of a certain sound - this would not account for the existence of a unified

experience, on the part of the conscious subject, of the book, hands, and table as a whole. As Hasker notes, if each student in a class knows the answer to at least one question in an examination, it doesn't follow that there is anyone who knows all the answers all at once. Their individual consciousnesses of the answer don't add up to a single, unified, collective consciousness of everything on the exam. Similarly, distinct neural processes correlated with different aspects of an object or scene by themselves do not, even if they are individually conscious, add up to consciousness of the object or scene as a whole (and things are only more mysterious when we keep in mind that these processes are not individually conscious.) Nor will positing the existence of some neural scanning mechanism along the lines of the higher-order states we've discussed in this chapter, which integrates the information in each distinct neural process, solve the problem. For now all the relevant information would have to be gathered together in this mechanism, which itself would be , composed of yet further distinct neural processes encoding distinct aspects of the visual field, and the binding problem would arise again at a higher level.

The implication seems to be that whatever it is that ultimately binds together the information presented either in perceptual experience or

*Consciousness* 111

in higher –order introspective awareness cannot be composed of parts which individually correlate with different aspects of the information. This would seem to lend some credence to Descartes's indivisibility argument, according to which the mind is a simple, and thus immaterial, substance. And it indicates that giving a materialist account of intentionality - which must ultimately be an account of the subject whose mind is directed upon an object when in an intentional state – is going to be a tall order indeed. Nevertheless, as we will see in the next two chapters, many materialists have tried to demonstrate that their view can meet this challenge.

### **Further reading**



The Block, Flanagan, and Guzeldere anthology *The Nature of Consciousness*, cited in the previous chapter, gives a large and representative sample of the enormous literature on consciousness that has developed over the last twenty years or so. Other important anthologies are Martin Davies and Glyn W. Humphreys, eds., *Consciousness* (Oxford: Blackwell, 1993), Thomas Metzinger, ed., *Conscious Experience* (Thorverton: Imprint Academic, 1995), and Quentin Smith and Aleksandar Jokic, eds., *Consciousness: New Philosophical Perspectives* (Oxford: Clarendon Press, 2003).

Dennett's eliminativism is defended in his influential book *Consciousness Explained*, cited in the previous chapter, and in "Quining Qualia," available in the Chalmers *Philosophy of Mind* anthology, also cited there. Cited there too was Chalmers' *The Conscious Mind*, in which he gives sympathetic treatments of both Russellianism and panpsychism. Other important book-length studies of the problem of consciousness include Owen Flanagan, *Consciousness Reconsidered* (Cambridge, MA: The MIT Press, 1992), William G. Lycan, *Consciousness* (Cambridge, MA: The MIT Press, 1987), and David Papineau, *Thinking About Consciousness* (Oxford: Oxford University Press, 2002), Joseph Levine's *Purple Haze: The Puzzle of Consciousness* (Oxford: Oxford University Press, 2001) is a rigorous critical analysis of all the most influential theories of consciousness, though the beginner will find it very hard going in places.

Representationalism is defended by Fred Dretske in *Naturalizing the Mind* (Cambridge, MA: The MIT Press, 1995), William G. Lycan in *Consciousness and Experience* (Cambridge, MA: The MIT Press, 1996), and Michael Tye in *Ten Problems of Consciousness* (Cambridge, MA: the MIT Press, 1995). Higher-Order theories are defended in Lycan's

## 112 *Philosophy of Mind*

*Consciousness and Experience*, D. M. Armstrong's "What Is Consciousness?" and David Rosenthal's "A Theory of Consciousness," the latter two essays being available in the Block, Flanagan, and Guzeldere anthology. Tim Crane's *Elements of Mind* (Oxford: Oxford University Press, 2001) contains his fullest exposition and defense of intentionalism.

Berkeley's *Principles of Human Knowledge* is available in many editions. An important contemporary defense of idealism is to be found in John Foster, *The Case for Idealism* (London: Routledge and Kegan Paul, 1982).

Russell's position is most fully developed in his *The Analysis of Matter* (London: Kegan Paul, 1927). (His remark about what the physiologist sees is on p. 383 of that book.) He briefly and lucidly summarizes it in chapter 2 of *My Philosophical Development* (London: Unwin Paperbacks, 1985). Recent defenders of the Russellian view include, in addition to Chalmers, Michael Lockwood, *Mind, Brain, and the Quantum* (Oxford: Basil Blackwell, 1989), Grover Maxwell, "Rigid Designators and Mind-Brain Identity," available in Chalmers' *Philosophy of Mind* anthology, and Galen Strawson, "Real Materialism," in Louise M. Antony and Norbert Hornstein, eds., *Chomsky and His Critics* (Oxford: Blackwell, 2003). Lockwood's book includes his defense of the notion of unsensed qualia (or phenomenal qualities, as he refers to them), a defense I criticize at greater length in "Can Phenomenal Qualities Exist Unperceived?", *Journal of Consciousness Studies* Vol. 5, No.4 (September 1998).

Searle develops the notion of aspectual shape, defends the connection principle, and criticizes materialist theories of consciousness in *The Rediscovery of the Mind* (Cambridge, MA: The MIT Press, 1992). Also of interest is Searle's *The Mystery of Consciousness* (New York: The New York Review of Books, 1997), which includes trenchant criticisms of, and testy exchanges with, Chalmers and Dennett. Hasker presents his argument from the unity of consciousness in *The Emergent Self* (Ithaca: Cornell University Press, 1999).