Consciousness

Tim Bayne
University of Oxford and St. Catherine's College
Manor Road
Oxford OX1 3UJ
United Kingdom
tim.bayne@gmail.com

Forthcoming in J. Symons & P. Calvo (eds) *Routledge Companion to the Philosophy of Psychology*. Routledge. This version of the paper is only a **draft**. For purposes of quotation please consult the published version.

After being sorely neglected for some time, consciousness is well and truly back on the philosophical and scientific agenda. This entry provides a whistle-stop tour of some recent debates surrounding consciousness, with a particular focus on issues relevant to the scientific study of consciousness. The first half of this entry (the first to fourth sections) focuses on clarifying the *explanandum* of a science of consciousness and identifying constraints on an adequate account of consciousness; the second half of this entry (the fifth to seventh sections) examines a number of the methodological challenges facing the science of consciousness.

The concept(s) of consciousness

Discussions of consciousness are plagued by disputes about how best to hone in on the phenomenon in question. How can we pre-theoretically characterize the object of study? Indeed, is there but one object of study here, or do multiple types of states go under the label "consciousness"?

Let's start with an easy distinction—or at least, a distinction that should be easy. Sometimes people (or organisms more generally) are said to be conscious, at other times mental states are said to be conscious. This distinction has come to be known as the distinction between creature consciousness and state consciousness. But although widely invoked, the distinction between creature consciousness and state consciousness is not always employed in the same way. Some theorists take creature consciousness to be the

property of being awake; on this view, creature consciousness is importantly distinct from state consciousness, and it is only state consciousness that is deeply puzzling. Other theorists see state consciousness and creature consciousness as intimately related to each other. According to one version of this view, creature consciousness is the determinable of which various states of consciousness are determinates: to say that a creature is conscious is to say that it is in some state of consciousness or other. This conception of creature consciousness allows that a creature could be conscious without be awake, for dream states qualify as states of consciousness. I will use "creature consciousness" in this latter sense, for it seems to me that to use "creature consciousness" to refer to the state of being awake merely fosters confusion.

Whether we are talking about conscious states or conscious creatures, we need a pre-theoretical gloss on consciousness if we are to get the project of theory-building off on the right foot. If there is any lesson to be learnt from the recent discussion of consciousness it is that any such gloss is very difficult to come by. Not only is there disagreement as to whether there is a single central form of consciousness, those who hold that there is a single core notion of consciousness often have very different conceptions of what that core notion is. Here, as elsewhere in philosophy, it can be difficult to tell whether the disagreements are substantive or terminological or both. We can get a fix on some of the central disputes in this domain by considering three notions with which consciousness is often associated: qualia, subjectivity, and phenomenality.

Qualia

In its most neutral sense, the term "qualia" refers to the appearances of things. The look of a tomato, the sound of a trumpet, the taste of a cucumber all involve qualia to the extent that these states represent how objects appear. In this thin sense of the term, no one would deny that there are qualia. However, many theorists employ one or another of various theoretically loaded senses of "qualia"—senses in which the very existence of qualia is up for grabs. Block, for example, uses "qualia" to refer to the non-intentional features of conscious states. One could deny that Block-qualia exist without denying the existence of consciousness—indeed, many do. Dennett (1988) also uses "qualia" in a theoretically loaded sense. Dennett-qualia, if they existed, would be ineffable, intrinsic, private, and directly apprehensible properties of mental states. Again, one could deny that Dennett-qualia exist without denying the existence of consciousness, as Dennett himself does. Even the claim that qualia must be conscious is not universally endorsed, with some theorists arguing in defence of non-conscious qualia.

Qualia are often associated with the so-called secondary qualities (colours, tastes, smells, and so on), but it is clear that primary properties—such as spatial extension, solidity, motion—also appear to us in consciousness. Are there qualia for these properties? Some say "yes," some say "no." Such debates are most plausibly regarded as terminological squabbles about how "qualia" is to be used rather than substantive disagreements about the scope of consciousness. Although theorists often describe themselves (and more often their opponents) as qualiaphiles or -phobes, in the absence of an unequivocal notion of qualia these terms threatens to generate more heat than light.

Subjectivity

According to many, the deepest problems of consciousness concern its supposed subjectivity (Levine 2001). Here too there is disagreement, both over what subjectivity itself involves and about how consciousness and subjectivity are related. A minimal conception of subjectivity identifies it with having a point of view. Most theorists regard consciousness as correlative with subjectivity in this minimal sense: to be conscious is to be a subject of experience, which in turn involves having a perspective on the world. A stronger conception of subjectivity ties the notion to self-consciousness, which is in turn unpacked in terms of having a conception of oneself as having a point of view or perspective. Just how this stronger notion of subjectivity might be related to consciousness is contentious. Some theorists would wish to sharply distinguish consciousness from strong subjectivity, holding that consciousness is one thing and self-consciousness is quite another. Other theorists argue that consciousness is constitutively dependent on strong subjectivity, and that one cannot be conscious without enjoying some form of self-consciousness.

Phenomenality

Arguably, the central notion of consciousness is that of phenomenality. Attempts to explicate the notion of phenomenality typically begin with examples: there is a distinctive phenomenal state associated with the taste of strawberries, there is another phenomenal state associated with the feeling of pain, and there is yet another phenomenal state associated with the smell of coffee. Each of these phenomenal states involves a distinctive phenomenal character – a distinctive "what-it's-likeness" (Nagel 1974). What it's like to taste strawberries differs from what it's like to experience pain, and both of these states differ in turn from what it's like to smell coffee. But insofar as each of these states is a phenomenal state it will possess the property of phenomenality: that is, there will be something it is like to instantiate it. Some theorists find what-it's-like talk illuminating, others find it less than helpful. The phrase is clearly not to be understood comparatively—

for although there is something that is similar to the taste of strawberries, this is clearly not at issue—but alternative analyses of the phrase are not easily had. The fact that it is so difficult to explicate the notion of what-it's-likeness raises the question of whether theorists who disagree about how to explain phenomenal consciousness have a shared conception of the notion, or whether they might (to some degree at least) be offering accounts of different features or forms of consciousness.

The notion of phenomenal consciousness is arguably the central notion of consciousness; at any rate, it is the notion on which I will focus here. In part, this is because it is the form of consciousness on which most philosophical ink has been spilled, but it is also because it is the form of consciousness that appears to be most resistant to scientific investigation. We have good reason to think that phenomenality is grounded in neural processes, but we lack any clear conception of how it is that phenomenality might be so grounded.

The contents of consciousness

What kinds of states can be phenomenally conscious? Or, to put the question in another way, what kinds of contents are phenomenally admissible? This question has an important bearing on the science of consciousness, for we might learn something (indeed, a lot) about what consciousness is by learning what kinds of mental states can (and cannot) be phenomenally conscious. Let us examine three domains in which there is disagreement about the "reach" of phenomenality. In each domain, we can distinguish a conservative position from a more liberal one.

Consider first vision. According to an oft-told story, visual experience (that is, visual phenomenology) can represent objects as being variously shaped, coloured, and standing in certain spatial and temporal relations to each other, but it cannot represent objects as belonging to particular scientific categories, or as being particular individuals, as having dispositional properties, and so on. Liberals challenge this traditional picture, arguing that such "high-level" properties—such as being a tractor, being a pine tree, or being Mrs Rosenblatt—can enter into the content of visual phenomenology (see e.g. Siegel 2006).

A second locus of debate between conservatives and liberals concerns the phenomenology of agency (Bayne 2007; Horgan et al. 2003). The experience of being an agent is ubiquitous: we experience ourselves as engaged in both physical activity (opening doors; pulling ropes) and mental activity (deliberating about a career decision; trying to pay attention to a dull speaker). How should we conceptualise such experiences? Conservatives argue that agentive experience is thin and austere—that its content is limited to (say) the experience of effort. Liberals argue that agentive experience is rich with content—that its content includes not only the experience of effort but also experiences of freedom, deliberation,

volition, and various kinds of causal relations. Of course, liberals will be quick to point out that just because we experience ourselves as (say) free or deliberating it does not follow that we are free or deliberating—in principle, for such experiences could be non-veridical.

From the perspective of the scientific study of consciousness, perhaps the most important domain in which the debate between conservatives and liberals is being played out concerns cognition. Is phenomenal consciousness restricted to perception, or does it permeate the cognitive realm? Liberals argue that there is a distinctive phenomenology associated with cognition (Pitt 2004; Siewert 1998; Strawson 1994). Some liberals hold only that the various propositional attitudes—intending, desiring, judging, and so on—have (or are associated with) proprietary phenomenal characters; others hold that particular propositional states, such as <intending to go to Paris> or <judging that camels have toes>, possess (or are associated with) distinctive phenomenal characters. Conservatives, by contrast, hold that phenomenality is the exclusive provenance of perception and sensation, and that there is no proprietary cognitive phenomenology no distinctive "what-it's-likeness" associated with either propositional attitudes or propositional attitude states. (Carruthers 2005; Lormand 1996; Tye 1995). According to the conservative, what it's like to (say) judge that camels have toes is exhausted by whatever bodily, imagistic, or perceptual states happen to accompany such a judgment.

Arguments for cognitive phenomenology typically involve appeals to phenomenal contrasts (Kriegel 2007). What it is like to hear the sentence "Il fait froid" when one does not understand French differs from what it is like to hear the same sentence having learnt French. In a similar vein, what it is like to hear the sentence "Visiting relatives can be boring" depends on whether one takes the "boring" to qualify the relatives or the visiting. In response to such arguments, typically grant that the difference between the two scenarios does involve "what-it's-like" differences, but they insist that such differences can be fully accounted for in sensory or perceptual terms. Needless to say, proponents of cognitive phenomenology remain unconvinced by such claims. They allow that differences in understanding and interpretation might (typically) involve differences in sensory experience, but they deny that such differences exhaust the phenomenal contrasts present in such cases.

Leaving aside the apparent phenomenology of cognition proper, conservatives must also reckon with the phenomenal fringe. By this I mean such states as feelings of knowing, tip-of-the-tongue experiences, déjà vu and jamais vu experiences, and similar meta-cognitive states. Such states count as cognitive in some sense of the term, and thus put pressure on

those forms of phenomenal conservatism that would restrict phenomenality to the purely sensory.

The debate between liberal and conservative conceptions of the scope of phenomenal consciousness is very puzzling. Why might it be so difficult to reach consensus on the reach of phenomenality? Leaving aside the possibility that conservatives and liberals are right about their own phenomenology but wrong only in generalizing from their own case to human phenomenology in general, there seem to be only two explanations of the debate: either introspection is leading one side of the debate seriously astray (Schwitzgebel 2008), or the debate is a terminological one, with conservatives and liberals talking passed each other. Whichever of these diagnoses is right—and perhaps they both contribute to the confusion—it is clear that the science of consciousness cannot afford to set the debate between conservatives and liberals to one side.

Phenomenality and intentionality

One of the many fundamental issues in consciousness studies concerns the relationship between phenomenal character (or phenomenality more generally) and intentional content (or intentionality more generally). How are these two aspects of the mind related? Are they fundamentally distinct, or are they deeply related in some way?

Although classical discussions of the mind from the medieval period right up to the phenomenologists tended to approach consciousness and intentionality in tandem, modern philosophy of mind has been dominated by what Horgan and Tienson (2003) have dubbed "separatism." According to the separatist, there is no internal relationship between phenomenal properties and intentional properties. The widespread endorsement of separatism has led many to adopt a "divide-and-conquer" approach to the problems of phenomenality and intentionality: theorists have assumed that accounts of phenomenality place few (if any) constraints on accounts of intentionality, and vice versa. This assumption would need to be re-evaluated were separatism to be rejected.

Separatism has a number of motivations. Perhaps most fundamentally, intentionality and phenomenality have seemed to many to be very different types of things. The gap between physical/functional properties and phenomenal properties has seemed to many to be deeper and more daunting than the gap between physical/functional properties and intentional properties. In a similar vein, phenomenality generates a problem of other minds that seems not to apply quite so forcefully to intentionality. We have some grip on how intentionality might be grounded in the natural world, but phenomenality seems to float free of any attempt to naturalize it. Whereas intentionality raises *problems*, phenomenality raises *puzzles*.

A less abstract—but perhaps more influential—argument for separatism concerns the possibility of dissociations between intentionality and phenomenality. According to many, propositional-attitudes states such as judgements and intentions have intentional content—they are about states of affairs—but there is nothing it is like to be in them. Consider the fact that a person can be said to believe that anchovies are fish or desire to win the Indy 500, even while they are in a dreamless sleep. Just how compelling one finds this argument depends on one's views of cognitive phenomenology. Proponents of cognitive phenomenology might grant that propositional attitudes have no phenomenal character when understood as dispositions, but they will insist that such states have phenomenal character when occurrent.

A further argument for separatism derives from information-processing models of linguistics and psychology. These models posit multiple levels of content-bearing states, none of which seem to possess any iota of phenomenal character. In response to this challenge, some inseparatists attempt to restrict the connection between intentionality and phenomenality to personal-level intentional states; others argue that the content-bearing states posited by information-processing psychology enjoy only ersatz or "as-if" intentionality, and that states with real intentional content must possess phenomenal character. Neither of these responses is entirely unproblematic.

Can there be mental states that have phenomenality without intentionality? At one point in the not-so-distant past an affirmative answer to this question would have been uncontroversial, and the student of consciousness would have been told that pains, moods, and orgasms (for example) enjoy phenomenal character but lack intentional content. Times have changed, and representational (or intentional) analyses of such states is now commonplace if not orthodox. The phenomenal character of moods is said to represent the general state of the world; the phenomenal character of pains is said to represent the relevant body part as damaged in some way or other; and the phenomenal character of orgasm is taken to represent "a certain change down there as good." Needless to say, opinions vary widely on the plausibility of such analyses.

Whether or not all states with intentionality have phenomenality (and vice versa), we can ask how the intentional content is related to the phenomenal character of those states that do enjoy both intentional content and phenomenal character. Does the intentional content of a state exhaust its phenomenal character, or vice versa? Could the phenomenal character (intentional content) of a state vary independently of its intentional content (phenomenal character), or does fixing the one also fix the other? Certain reductive projects assume that phenomenal character (intentional content) cannot vary independently of intentional content (phenomenal character). In recent times, the most

common reductive project in this vicinity has been to reduce phenomenal character to a certain kind of intentional content, but one might also attempt to reduce intentional content to a certain kind of phenomenal character.

Such reductive projects must reckon with inverted-spectrum (or inverted-qualia) scenarios. Some inverted-spectrum scenarios are designed to elicit the intuition that phenomenal character can vary independently of intentional content, whereas others are designed to elicit the intuition that intentional content can vary independently of phenomenal character. Inverted-spectra scenarios typically turn on the fact that intentional content seems to be "wide"—that is, dependent on the subject's environment and/or history—whereas phenomenal character seems to be "narrow," that is, independent of the subject's environment and/or history. The apparent wideness of intentional content allows it to vary independently of variations in phenomenal character.

Reductionists have three options open to them in responding to inverted-spectrum scenarios. Some theorists—"phenomenal externalists"—reject the assumption that phenomenal character is narrow. They hold that the phenomenal character of a state is determined by its environmental and historical relations. Other theorists reject the assumption that intentional content is wide. They hold that the intentional content of a state is determined by the intrinsic properties of the subject in question; indeed, perhaps those very same properties that determine the subject's phenomenal states A third group of theorists attempt to finesse the problem by distinguishing between narrow content, which is essentially tied to phenomenal character, and wide content, which is not. Each of these three positions has its advocates, and the precise relation between phenomenal character and intentional content is likely to remain on the philosophical agenda for some time (Chalmers 2004).

The structure of consciousness

At any one time, a subject's various experiences occur as parts (components, aspects) of a total phenomenal state. This state subsumes each of the experiences that the subject has at the time in question—it captures what it is like to be that subject. This is one fairly natural way to understand the claim that consciousness is unified (see also Dainton 2005; Hurley 1998; Tye 2003). Is consciousness *necessarily* unified, or is it possible for the unity of consciousness to break down under certain conditions? In other words, is it possible that a subject might have simultaneous experiences that are not subsumed by a single phenomenal state? The answer to this question bears on the scientific study of consciousness, for an account of consciousness ought to explain why consciousness is

necessarily unified if it is, and it ought to explain why the unity of consciousness can break down if indeed it can.

Among the most striking of those conditions in which the unity of consciousness appears to break down is the commissurotomy (or "split-brain") syndrome. This syndrome results from a procedure in which the corpus callosum—the bundle of fibres that connect the two hemispheres of the brain—is severed in an effort to prevent epileptic seizures from migrating between hemispheres. Although split-brain patients show few signs of disunity in everyday life, under carefully controlled laboratory conditions they can be led to behave in ways that suggest that their consciousness is no longer unified. For example, if the word "key-ring" is presented so that the words "key" and "ring" fall within the patient's left and right visual hemi-fields respectively, the patient will say that she sees only the word "ring" when asked to report what she sees, yet she may select a picture of a key with her left hand while ignoring pictures of a ring and a key-ring.

According to many theorists, such behaviour indicates that split-brain patients have two streams of consciousness, one in each hemisphere. It is this duality of consciousness that is thought to explain (i) why the patient appears to have conscious representations of "key" and "ring" but no representation of "key-ring"; and (ii) why the patient's representations of "key" and "ring" are available to different consuming systems. Another model of the split-brain, suggested by Michael Lockwood (1989), takes split-brain patients to have a fragmented or partially unified stream of consciousness, with triplets of conscious states (e_1, e_2, and e_3) such that e_1 and e_2 are both unified with e_3 but not with each other.

Both the two-streams and partial-unity models entail that split-brain patients lack a unified consciousness. A third account of the split-brain—the switch model—holds that consciousness remains unified even in the split-brain. According to this model, the appearance of simultaneous but distinct streams of consciousness in the split-brain patient is an illusion generated by the fact that consciousness switches rapidly and seamlessly between the patient's two hemispheres; at no time does the patient have simultaneously conscious states that are not phenomenally unified (Bayne 2008). The split-brain patient may have conscious representations of the word "key" and "ring," but these representations will be sequential rather than simultaneous.

The debate between these three accounts raises a number of interconnected issues. Two-streams accounts of the split-brain are well-equipped to explain the behavioural disunity that patients exhibit in laboratory conditions, but they struggle to account for the unity that patients exhibit in everyday situations. Lockwood's partial-unity model might do better at explaining the behaviour of split-brain patients in both everyday and

experimental settings, but it is unclear whether consciousness can be partially unified. The switch model promises to explain the unity of the patient's behaviour and everyday life and the disunity of their laboratory behaviour, but faces the challenge of explaining why patients fail to realize that their stream of consciousness shuttles between hemispheres.

Whether or not the unity of consciousness might break down in the context of the split-brain syndrome (or other pathologies of consciousness), it is clearly unified in everyday life. Broadly speaking, there are two ways in which to conceptualise this unity: in terms of experiential building blocks or in terms of a unified phenomenal field (Searle 2000). The building-block approach regards consciousness as built up out of atomistic phenomenal states that are put together to form the subject's total phenomenal state. By contrast, the unified-field theorist regards consciousness as having a field-based structure, according to which subjects enjoy particular phenomenal states—the state of hearing a siren, having a headache, smelling strawberries—only in virtue of contents of the appropriate kind feeding into a phenomenal field. The unified-field approach to consciousness does not do away with the problem of accounting for the unity of consciousness, but it does perhaps make the problem more tractable.

There is some reason to prefer the unified-field model over the building-block model. If consciousness had a building-block structure, then we would expect there to be pathologies of consciousness in which consciousness would fragment into its constituent elements. However, there appear to be no such pathologies. To be sure, there are pathologies of consciousness in which the normal representational coherence and integration of consciousness is lost. For example, patients with visual agnosia can see the various parts of an object, but they cannot synthesize those components so as to form the representation of a unified object. But there is no reason to suppose that agnosics have visual experiences that fail to be phenomenally unified. Arguably, the pathology of consciousness in which the unity of consciousness seems most likely to break down is the split-brain syndrome, but we have seen that even here the case against the unity of consciousness is far from decisive. And even if consciousness is disunified in the split-brain, the units of consciousness that the split-brain patient enjoys are not plausibly regarded as the basic units of which a subject's total conscious state might normally be built.

Criteria for the ascription of consciousness

In order to study a phenomenon one needs to know which objects possess it and which objects do not. One need not be infallible when it comes to identifying the target property, but the poorer one's methods of detecting a property the harder it is to develop a theory

of it. In light of this, the science of consciousness stands in need of criteria for the ascription of consciousness. We need decision-procedures by means of which we can tell whether very young human beings, the comatose, non-human animals and robots might be conscious.

Many scientists hold that the only reliable measure of consciousness is introspective report: to be justified in thinking that a subject was (or was not) conscious of a certain stimulus one must have elicited an appropriate introspective report from him or her. Accepting the introspective-report criterion has important ramifications for the scientific study of consciousness, for it radically curtails the subject pool from which such studies can draw. Young children, non-linguistic animals, and those in abnormal states of consciousness (such as delirium or hypnosis) are typically unable to produce introspective reports, yet it is often precisely such subjects that we wish to study (especially if we want to model non-standard forms of consciousness). The scientific study of consciousness would be that much more difficult were we to restrict ourselves to introspective reports.

Luckily, there is good reason not to restrict ourselves to introspective reports. Firstly, despite their official pronouncements, few studies of consciousness actually rely on introspective reports. Instead, studies typically ask subjects to report on the nature of their environment. Subjects are asked to respond whenever they (say) detect a red light, not whenever they detect an *experience* as of a red light. Of course, proponents of the introspective-report criterion might argue that environmental reports are operationally equivalent to introspective reports, at least where normal adult human beings are concerned. It is not obvious that this is so, but even if it is, most consciousness scientists are willing to accept the behavioural responses of creatures that are incapable of giving introspective reports. In their influential work using a binocular-rivalry paradigm in monkeys, Logothetis and colleagues trained monkeys to press levers depending on which of two pictures were presented to them (Logothetis and Schall 1989). Although I am inclined to grant that this behaviour can be used to track the conscious states of the monkeys, it is not even clear to me that the monkeys were producing reports of any kind, let alone introspective reports.

Before we turn to alternative criteria for the ascription of consciousness, it is worth asking why theorists emphasize the importance of introspective reports in studying consciousness. One motivation for the view may derive from the assumption that introspection is infallible, or at least highly reliable. In light of this, it is somewhat ironic that many of those who place most weight on the importance of introspective reports for the study of consciousness also hold that introspective judgments can be massively mistaken. They hold that we are subject to a "grand illusion" when it comes to perceptual

consciousness—we ordinarily think (that is, believe) that we are conscious of much more than we are in fact conscious of. Evidence in favour of the grand illusion is controversial, but there is at least some reason to think that subjects' introspective judgments can be erroneous. For example, it is commonly believed that colour experience extends to the periphery of one's visual field, but it is at least arguable that we do not experience peripheral objects as coloured. But note that if we were serious about taking (sincere) introspective reports as the "gold standard" of consciousness, then we could never have good reason for thinking that introspection can lead subjects astray. Since we do have reason to think that introspection can lead subjects astray, we should refuse to take introspective reports as our sole criterion for the ascription of consciousness.

Another worry with relying on introspective reports to study consciousness is that subjects might fail to notice (and hence report) that they are in certain phenomenal states. The question of whether consciousness might outrun reportability is of more than merely of theoretical interest. Consider the perceptual extinction paradigm, employed by Rees and colleagues for studying the neural correlates of visual consciousness (Rees et al. 2002). Perceptual extinction is a phenomenon in which stroke patients ignore stimuli presented in one visual field (typically the left) when such stimuli are presented together with stimuli in the other visual field. Rees and colleagues assume that patient are not conscious of the ignored (that is, unreported) stimuli, but one might argue that such stimuli are conscious but merely unnoticed.

Let us return to the question of what criteria we should adopt for the ascription of consciousness. One proposal worth serious consideration is that the capacity for intentional agency suffices to underwrite the ascription of consciousness. Consider the following functional magnetic resonance imaging study of a woman in a vegetative state (Owen et al. 2006). The patient was scanned while being asked to first imagine visiting the rooms in her house and then imagine playing tennis. The patterns of brain activation observed suggested that she had engaged in the appropriate mental imagery: the request to imagine visiting the rooms of her house generated activity in the parahippocampal gyrus, posterior parietal cortex, and lateral premotor cortex, whereas the request to imagine playing tennis generated activity in the supplementary motor areas that control motor responses. Does this activation provide good evidence that the patient was conscious? There is some temptation to think that it does. Arguably, the link between the neural activity observed here and the ascription of consciousness goes via the notion of intentional agency.

The criteria for the ascription of consciousness discussed thus far have been exclusively functional, but on might have reservations about purely functionalist approaches to the

ascription of consciousness. For one thing, one might worry about zombies—that is, unconscious creatures whose functional profile mirrors our own. A purely functional approach to the problem of other minds would lead one to (falsely) regard one's zombie twin as conscious. Of course, it is controversial whether zombies are possible, but it's not obvious that they are impossible. Furthermore, even if *pure* zombies are impossible, partial zombies—that is, unconscious creatures that depart from our own functional profile in only minor ways—may well be possible, and a functionalist approach to the ascription of consciousness might lead one to regard one's partial-zombie twin as conscious.

Functional measures of consciousness also run the risk of excluding some of those who actually are conscious. Locked-in syndrome patients suffer from almost total paralysis and can communicate only by moving their eyes and/or blinking. Despite their behavioural incapacitation, locked-in patients are fully conscious; indeed, they can *tell* us that they are conscious. But now imagine a locked-in patient who loses even the ability to control the movements of his or her eyes but retains consciousness. A purely functional approach to the problem of other phenomenal minds might (falsely) pronounce such a person unconscious. Clearly we should be reluctant to employ any criteria for the ascription of consciousness that would lead us to regard a conscious human being as unconscious.

There is no straightforward solution to this problem. We could replace (or supplement) functional criteria for the ascription of consciousness with physical criteria. Perhaps the stuff of which an entity is made has a direct bearing—that is, a bearing independently of the functional roles that it realizes—on whether or not it is conscious. But this approach to the ascription of consciousness can easily be made to look unattractive. Suppose that you and I are made of different types of materials: I turn out to be carbon based whereas you turn out to be silicon based. Or suppose that we discover that I have only a left cerebral hemisphere whereas you have only a right cerebral hemisphere. Why should this (or similar) information have an impact on whether it is reasonable for me to regard you as conscious (or vice versa)? Surely it would be chauvinistic to suppose that an entity must be built out of the same stuff as us (or should I say "me"?) in order to be conscious.

In response to all of this, one might argue that even if our present methods for detecting consciousness are crude, confused, or just plain wrong, this state of affairs is not intractable: as our understanding of consciousness improves, so too will our ability to determine whether or not a creature is consciousness. We might be currently unsure as to how to weigh physical and functional properties in ascribing consciousness, but our difficulties here may be merely temporary (Block 2007).

But how exactly are we to develop scientifically grounded criteria for the ascription of consciousness? One strategy would be to start with paradigm cases of consciousness and take as one's criteria for consciousness whatever physical or functional properties apply in such cases. The problem with this approach is that only some of the physical and functional properties that apply in the paradigm cases will apply in the hard cases, and we will need to decide whether the similarities between the paradigm case and the hard case are strong enough to justify ascribing consciousness in the hard cases (Block 2002). In other words, the paradigm-case approach appears to be impotent exactly where it is most needed. An alternative strategy appeals to theories of consciousness. If we knew where in the architecture of cognition conscious states are generated, or on what kinds of physical or functional states consciousness supervenes, then we could tell what kinds of conscious states, if any, a target organism instantiates. The problem with this strategy is that in order to develop a theory of consciousness one needs to have criteria for the ascription of consciousness in place, and that is precisely what we don't have. Arguably, to develop rigorous criteria for the ascription of consciousness we would need to close – or at least narrow – the explanatory gap. It is only when we have a grasp of how functional and/or physical states realize or generate phenomenal states that we will be in a position to determine with any certainty whether or not organisms that differ radically from ourselves are conscious. There may be a degree of intractability about the distribution of consciousness if, as many suspect, the explanatory gap cannot be closed.

Content-based vs. creature-based methodologies

Let us leave to one side problems concerned with the ascription of consciousness and focus instead on how the science of consciousness might proceed. Two explanatory strategies can be discerned in current attempts to find the mechanisms underpinning consciousness. We might call these two approaches the "content-based approach" and the "creature-based approach." The content-based approach focuses on the mechanisms underlying specific contents of consciousness. The experimenter contrasts one type of conscious state with another and looks for the neural changes that are most closely correlated with changes in the contents of consciousness. A classic example of this paradigm can be found in the binocular-rivalry studies of Logothetis and colleagues, mentioned earlier. Binocular rivalry occurs when different stimuli are presented to the left and right eyes, as the subject's visual experience usually alternates between the two stimuli. Logothetis and colleagues trained monkeys to press a bar, depending on which of two visual stimuli they were seeing, and then recorded the neural responses of monkeys in this paradigm. They found that activity in the primary visual cortex correlated strongly with changes in stimuli but only weakly with changes in the monkeys' visual experience,

whereas activity in later visual areas (such as inferior temporal cortex) correlated strongly with changes in visual experience. Such findings tempt many to think of later visual areas as the neural correlates of visual consciousness.

What this view overlooks is the fact that activity in the later visual areas generates visual experiences only when various domain-general systems are active. Proponents of the creature-based approach to the study of consciousness focus on the search for these mechanisms. Adopting a creature-based methodology, Laureys and others have examined the neural mechanisms underlying the transition between the comatose state, in which the patient is assumed to be unconscious, and the minimally conscious state, in which the patient is assumed to go in and out of consciousness (Laureys 2005).

There is at present no consensus concerning which of these two methodologies ought to be adopted. Those who privilege the content-based approach to the study of consciousness don't deny that there are domain-general mechanisms of creature consciousness, but they tend to downplay their importance, referring to them as merely "enabling conditions." By contrast, those who privilege the creature-based approach tend to downplay the relevance of those mechanisms that serve to differentiate one state of consciousness from another in explaining consciousness. How, they ask, could studying the contrast between one state of consciousness and another shed light on consciousness itself? In part, the debate between these two methodologies turns on a more fundamental debate about the structure of consciousness. As discussed earlier, some theorists take consciousness to be built up out of phenomenal building blocks, whereas others hold that consciousness has a holistic, field-based structure. Proponents of the building-block conception of the structure of consciousness are likely to endorse the content-based methodology, for by looking at the neural mechanisms that are uniquely necessary for certain conscious states one might discover particular phenomenal building blocks. By contrast, proponents of the field-based view of consciousness are likely to be more sympathetic to the creature-based methodology, for the creature-based methodology would seem best-suited for identifying the mechanisms underlying the generation of the phenomenal field. However, there is some sense to be made of content-based methodology, even if consciousness has a field-like structure. Unlike the building-block theorist, the phenomenal-field theorist will be more interested in the domain-general mechanisms that are implicated in all conscious states than in the content-specific cortical nodes that underpin the contrast between one conscious state and another. The jury is still out on which of these two approaches to the study of consciousness ought to be pursued; arguably, both approaches have a place within the science of consciousness.

Functional reduction and levels of explanation

What—if any—kind of explanation of consciousness might we hope to discover? Arguably the deepest divide in this domain is between those who think that consciousness will succumb to a reductive analysis in much the way that other natural phenomena such as—life, photosynthesis, reproduction, and so on—have succumbed, and those who hold that consciousness will resist such an analysis. Levine (1983), Chalmers (1996), and McGinn (1989), amongst others, argue that consciousness poses a unique explanatory gap, and that no amount of physical or functional knowledge would enable us to grasp how consciousness "arises out of" physical processes. Although these theorists deny that a deep or reductive explanation of consciousness will be found, they do not (or at least need not) deny that it might be possible to discover correlations between phenomenal and physical/functional properties. Finding these correlations and the mechanisms underlying them would provide us with an explanation of some kind, it's just that the explanation would leave something out.

The debate between these two camps has been waged on two fronts. One front concerns the prospects of an *a priori* functionalist analysis of phenomenal consciousness, and the implications that the failure of such an analysis might have for theories of consciousness. One line of thought runs as follows: To reduce a phenomenon one must first give an *a priori* functionalist analysis of it; since phenomenal consciousness cannot be so analysed it cannot be reduced. Some theorists take issue with the first claim, holding that phenomenal consciousness can be given an *a priori* functional analysis; others take issue with the second claim, arguing that a property can be reduced even if it cannot be given a functional analysis.

The second front in this debate has focused on why phenomenal consciousness seems to be uniquely resistant to scientific analysis. Whereas reductionist argue that phenomenal consciousness seems to be resistant because it is resistant, reductionists attempt to explain away that apparent refractoriness of consciousness to scientific explanation. A popular strategy here appeals to phenomenal concepts. Many of those who argue that consciousness can be given a reductive analysis argue that the sense of a unique explanatory gap between phenomenal properties and physical/functional properties is generated by the fact that our perspective on our own phenomenal states is mediated by phenomenal concepts—concepts that can be deployed only from the first-person perspective. According to this line, the sense that there is an explanatory gap between phenomenal properties and physical/functional properties is a cognitive illusion, generated by the radical difference between first-person and third-person access to phenomenal properties (Tye 1999; Papineau 2002).

As we have seen, questions of whether the science of consciousness might be able to deliver a reductive explanation of consciousness quickly descend into complex debates concerning the nature of explanation and the relationship between concepts and properties. These issues stand at some distance from the science of consciousness itself. Of more concern to the science of consciousness are questions about the level at which an account of consciousness ought to be pitched. Of course, should not assume that explanations of consciousness will be restricted to a single level of explanation, nonetheless, we can still usefully ask which explanatory level(s) will cast the most illumination on consciousness.

Some theorists appeal to the categories of folk psychology to explain consciousness. This is the approach that tends to be taken by proponents of higher-order theories of consciousness, who argue that a state is conscious in virtue of being the object of an appropriate higher-order thought or perception (Lycan 1996; Rosenthal 2004). (Note that such accounts are not always advanced as accounts of phenomenal consciousness.) According to some versions of the higher-order approach, states are conscious in virtue of being the intentional object of another mental state; according to other versions of the approach, states are conscious in virtue of taking themselves as their own intentional objects (Kriegel and Williford 2006). Either way, the higher-order account of consciousness invokes only those states and relations recognized by folk psychology. Functionalist-cum- representational theories of consciousness typically employ folk-psychological categories in accounting for consciousness. For example, Tye (1995) holds that a mental state is conscious in virtue of being poised for employment by the subject's conceptual system. Common to all of these theories is the assumption that although scientific progress may be needed in order to fill in the details of how consciousness is implemented in a particular species, the heavy lifting in an account of consciousness is done by such folk-psychological categories as higher-order representation, availability to conceptual deployment, and the like.

Other theorists regard advances in psychology and neuroscience as indispensable for understanding consciousness. Typically, those who hold this view begin by identifying a functional role that is associated with consciousness, and then look for the cognitive and neural mechanisms which realize that role. The gamma oscillations account of consciousness provides a useful example of this approach (see e.g. Crick and Koch 1990). Proponents of this approach argue that consciousness involves gamma oscillations in the sensory cortices—so-called 40-hertz (Hz) oscillations—on the grounds that such oscillations are thought to play a role in dynamic feature binding, the process by means of

which various perceptual features (representations of colour, shape, spatial location and orientation) – are bound together to form percepts of unified objects.

The 40-Hz proposal can be challenged on a number of grounds. Most fundamentally, it is unclear how robust the link between consciousness and feature binding is. What are the features that must be bound together in order to generate conscious moods, desires, intentions, and thoughts? Even where a mental state does require feature binding, it is not at all obvious that feature binding is sufficient for consciousness for there is some reason to think that binding can occur unconsciously. And even if one were to accept the connection between feature binding and consciousness, it is not obvious what implications that has for the connection between 40-Hz oscillations and consciousness. Proponents of the 40-Hz account of consciousness often suggest that consciousness might be *identified* with gamma oscillations but justifying an identity claim of this kind is problematic. Rather than identify consciousness with the mechanisms that realize feature binding in us, one could identify it with feature binding itself. Feature binding might be subserved by 40-Hz oscillations in humans but by an entirely different mechanism in other species. The contrast between these two positions has implications for the distribution of consciousness. Identifying consciousness itself with 40 Hz oscillations commits us to the view that there is nothing it is like to be a creature in which feature binding is realized by some other mechanism, whereas identifying consciousness with feature binding itself would allow us to regard such creatures as conscious.

It is now more common for cognitive neuroscientific accounts of consciousness to take global availability rather than feature binding as their starting point (Baars 1988; Dehaene and Naccache 2001). Global workspace accounts of consciousness identify consciousness with—a domain-general working memory system, the contents of which are available to a wide range of the subject's consuming systems. On some versions of the account, the global workspace has a fixed neural architecture (typically thought to involve parietal and prefrontal networks); other versions allow that the global workspace might involve dynamic coalitions of widely distributed cortical areas.

Global workspace approaches to consciousness are well motivated, for the contents of consciousness are typically available to a wide range of consuming systems (verbal report, the voluntary allocation of attention, memory consolidation, belief formation, and so on). Nonetheless, the approach is vulnerable to a number of challenges. For one thing, it is debatable whether there is a single global workspace as opposed to a number of less-than-global workspaces. The fact that the contents of consciousness are typically available to a wide range of consuming systems might result from the operation of a number of domain-specific working memory systems rather than a single executive system. Further,

it is not clear that all conscious contents make it into a working memory of any kind, let alone a domain-general one. One reason for distinguishing the contents of consciousness from those of working memory is that the two faculties appear to have different capacity constraints: on most accounts, working memory has at most a capacity of four items, whereas the capacity of phenomenal consciousness seems to be substantially larger. Arguably, many conscious states never make it from perceptual buffers into the global workspace (Block 2007).

Even if we were to associate consciousness with a global workspace, we would face the further question of whether to identify consciousness with the neural mechanisms that realize this workspace in us. Such a move would entail that creatures in which the global workspace is realized by some other neural or cognitive architecture would be unconscious, and it is by no means obvious that we ought to embrace this result. Understanding the mechanisms that allow the contents of our mental states to be globally available for cognitive and behavioural consumption will no doubt tell us something about consciousness, but it is unclear how much light it will shed on its essential nature.

Some theorists think that the fundamental explanation of consciousness will be found at the level of fundamental physics. Within this grouping some look to quantum theory; others, to some as-yet-undeveloped theory. Although theorists have various reasons for looking to physics for an account of consciousness, a common motivation for thinking that consciousness can be usefully tackled at a physical level involves the notion that phenomenal properties are not functional, and that it is only at the level of fundamental physics that we find non-functional properties. Many find this proposal implausible, for it appears to lead to some form of panpsychism: given that all objects presumably have these fundamental properties, it seems to follow from this view that all objects (and all parts of all objects!) are phenomenally conscious. Whatever attractions the physics-based approach might otherwise have, this implication alone suffices to rule it out of serious contention.

Conclusion

In *Empiricism and the Philosophy of Mind* Wilfrid Sellars distinguished two images of the world: the manifest image and the scientific image. Consciousness is a nonnegotiable component of the manifest image. From the first-person perspective, eliminativism is simply not an option. We cannot think phenomenality away, no matter how hard we try. But matters are very different from the third-person perspective of the scientific image. Consciousness is now taken as a serious subject of research within the cognitive sciences, but this should not obscure the fact that its scientific credentials are far from secure. It is, I

think, very much an open question whether the sciences of the mind *need* to invoke consciousness, or whether they will get along just fine with it. It certainly cannot be said that consciousness has already earned its explanatory keep. The notion of consciousness may be one of the few notions that we retain not because it pays its way but because we are stuck with it.

References

Baars, B. J. (1988) *A Cognitive Theory of Consciousness*, Cambridge: Cambridge University Press.

Bayne, T. (2008a) "The Phenomenology of Agency," *Philosophy Compass*, 3: 1–21.

——— (2008b) "The Unity of Consciousness and the Split-Brain Syndrome," *Journal of Philosophy*,

Block, N. (2002) "The Harder Problem of Consciousness," *Journal of Philosophy* 99: 391–425.

——— (2007) "The Methodological Puzzle of Consciousness and the Mesh between Psychology and Neuroscience," *Behavioral and Brain Science* 30: 481–548.

Carruthers, P. (2005) "Conscious Experience versus Conscious Thought," in *Consciousness: Essays from a Higher-Order Perspective*, Oxford: Oxford University Press, pp. 134–56.

Chalmers, D. (1996) *The Conscious Mind*, Oxford: Oxford University Press.

——— (2004) "The Representational Character of Experience," in B. Leiter (ed.) *The Future for Philosophy*, Oxford: Oxford University Press, pp. 153–81.

Crick, F., and Koch, C. (1990) "Towards a Neurobiological Theory of Consciousness," *Seminars in the Neurosciences* 2: 263–75.

Dainton, B. (2005) *Stream of Consciousness*, 2nd edn, London: Routledge.

Dehaene, S., and Naccache, L. (2001) "Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework," *Cognition* 79: 1–37.

Dennett, D. (1988) "Quining Qualia," in A. Marcel and E. Bisiach (eds), *Consciousness in Contemporary Science*, Oxford: Oxford University Press, pp. 42–77.

Dretske, F. (1995) *Naturalizing the mind*, Cambridge, MA: MIT Press.

Horgan, T., and Tienson, J. (2002) "The Phenomenology of Intentionality and the Intentionality of Phenomenology," in D. Chalmers (ed.) *Philosophy of Mind: Classical and Contemporary Readings*, Oxford: Oxford University Press, pp. 520–33.

Horgan, T., Tienson, J., and Graham, G. 2003. "The Phenomenology of First-Person Agency," in S. Walter and H.-D. Heckmann (eds) *Physicalism and Mental Causation: The Metaphysics of Mind and Action*, Exeter, UK: Imprint Academic, pp. 323–40.

Hurley, S. (1998) *Consciousness in Action*, Cambridge, MA: Harvard University Press.

Kriegel, U. (2007) "The Phenomenologically Manifest," *Phenomenology and the Cognitive Sciences* 6: 115–36.

Kriegel, U., and Williford, K. (eds) (2006) *Self-Representational Approaches to Consciousness*, Cambridge, MA: MIT Press.

Laureys, S. (2005) "The Neural Correlate of (Un)Awareness: Lessons from the Vegetative State," *Trends in Cognitive Sciences* 9, no. 12: 556–9.

Levine, J. (1983) "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly* 64: 354–61.

——— (2001) *Purple Haze*, Oxford: Oxford University Press.

Lockwood, M. (1989) *Mind, Brain and the Quantum*, Oxford: Blackwell.

Logothetis, N., and Schall, J. (1989) "Neuronal Correlates of Subjective Visual Perception," *Science* 245: 761–3.

Lormand, E. (1996) "Nonphenomenal Consciousness," *Nous* 30: 242–61.

Lycan, W. (1996) *Consciousnes and Experience*, Cambridge, MA: MIT Press.

McGinn, C. (1989) "Can We Solve the Mind-Body Problem?" *Mind* 98, no. 391: 349–66.

Nagel, T. (1974) "What Is It Like to Be a Bat?" *Philosophical Review* 83: 435–50.

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., and Pickard, J. D. (2006) "Detecting Awareness in the Vegetative State," *Science* 313 (8 September): 1402.

Pitt, D. (2004) "The Phenomenology of Cognition *or* What Is It Like to Think That *P*?" *Philosophy and Phenomenological Research* 69: 1–36.

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

Rees, G., Kreiman, G., and Koch, C. (2002) "Neural Correlates of Conscious and Unconscious Vision in Parietal Extinction," *Neuroscase* 8: 387–93.

Rosenthal, D. (2004) "Varieties of Higher-Order Theory," in R. Gennaro (ed.) *Higher-Order Theories of Consciousness*, Amsterdam and Philadelphia: John Benjamins, pp. 17–44.

Schwitzgebel, E. (2008) "The Unreliability of Naïve Introspection," *The Philosophical Review* 117, no. 2: 245–73.

Searle, J. (2000) “Consciousness,” *Annual Review of Neuroscience* 23: 557–78.

Siegel, S. (2006) “Which Properties Are Represented in Perception?” in T. Gendler and J. Hawthorne (eds) *Perceptual Experience*, Oxford: Oxford University Press, pp. 481–503.

Siewert, C. (1998) *The Significance of Consciousness*, Princeton, NJ: Princeton University Press.

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

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

——— (1999) “Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion,” *Mind* 108: 705–25.

——— (2003) *Consciousness and Persons*, Cambridge, MA: MIT Press.