

What is consciousness?

David Carmel
University of Edinburgh

Mark Sprevak
University of Edinburgh

6 April 2016

Human consciousness is one of the greatest mysteries in the universe. From one point of view this should be surprising, since we know a great deal about consciousness from our own experience. One could say that our own conscious experience is the thing in the world that we know best. Descartes wanted to build the entirety of natural science on the foundation of our understanding of our conscious thought. Yet despite our intimate relationship with our own consciousness experience, from another point of view consciousness is a puzzling phenomenon. We have no idea what it is about us, as physical beings, that makes us conscious, why we have consciousness, or which creatures other than humans have consciousness. Not only is it hard to answer these questions, it is hard to know how to even start to find answers.

1 The question of consciousness: Philosophical perspectives

Human consciousness is one of the greatest mysteries in the universe. From one point of view this should be surprising, since we know a great deal about consciousness from our own experience. One could say that our own conscious experience is the thing in the world that we know best. Descartes wanted to build the entirety of natural science on the foundation of our understanding of our conscious thought. Yet despite our intimate relationship with our own consciousness experience, from another point of view consciousness is a puzzling phenomenon. We have no idea what it is about us, as physical beings, that makes us conscious, why we have consciousness, or which creatures other than humans have consciousness. Not only is it hard to answer these questions, it is hard to know how to even start to find answers.

2 What do we talk about when we talk about consciousness?

We talk about consciousness in our everyday lives. We say that ‘she wasn’t conscious of the passing pedestrian’, that ‘he was knocked unconscious in the boxing ring’, that our ‘conscious experience’ of smelling a rose, making love, or hearing a symphony makes life worth living. Consciousness is what philosophers call a folk concept: a notion that has its home in, and is ingrained into, our everyday talk and interests. One problem that we encounter when trying to understand folk concepts is that they tend to be messy; they collect together diverse things that interest us under a single heading. When we investigate the world systematically we may start out with folk concepts, but we are often forced to refine or abandon them in favour of more precise scientific counterparts. Physics was forced to abandon the folk notions of *heaviness* and *speed* in favour of the concepts of *mass* and *velocity*, which allowed us to describe universal laws and build scientific theories. A science of *heaviness* or *fastness* would have been impossible because these folk notions collect too many diverse things under a single heading. A scientific understanding of consciousness, therefore, should approach our folk notion of consciousness with care. Although we use the words ‘conscious’ and ‘consciousness’ already, we might be using them to refer to a variety of different things, and we should distinguish between them. So what might we mean by ‘consciousness’?

One thing we might mean is sentience. When we say a creature is conscious of its surroundings, we mean that it is receptive to those surroundings and it can act in an intelligent way. For example, we might say that the spider under the fridge is conscious of our presence: the spider is sensitive to our presence and has sensibly taken evasive action. On this conception of consciousness, there is no difficulty with a robot or an amoeba being conscious; it simply means that the entity responds in a reasonable way to its environment.

A second, and distinct, meaning of ‘consciousness’ is wakefulness. When we say that someone is conscious, what we mean is that she is alert and awake: she is not asleep or incapacitated. For example, we might say that we were unconscious in dreamless sleep, or when knocked out by a blow to the head. This conception of consciousness suggests that consciousness is a global state, a kind of switch, which colours the whole mental life of a creature.

A third thing we might mean by ‘consciousness’ is higher-order consciousness. A creature has higher-order consciousness if it is aware of itself as a thinking subject. This requires not just that a creature have thoughts, but also that it be aware of – be capable of reflecting on – those thoughts. For example, a creature may not just think that it is too hot and act appropriately (take off clothes, seek cooler surroundings), it may also think that having its ‘I-am-too-hot’ thought is surprising

given the wintry weather. ‘Perhaps’, such a creature may think, ‘I am sick, and my “hot” thought has occurred because I am feverish?’ Such a creature doesn’t merely think about and perceive, it also thinks about its thoughts and perceptions. Higher-order consciousness requires *metacognition*: that a creature reflects on, and thinks thoughts about, its thoughts and perceptions.

A fourth thing we might mean by ‘consciousness’ is what the philosopher Ned Block (1995) terms access consciousness. A creature’s thought is access conscious if it is ready to interact with a wide variety of the creature’s other thoughts. A thought is access conscious if it is ‘broadcast widely’ in a creature’s brain. For humans, thoughts and perceptions that can be verbally reported are usually access conscious. Not all of your perceptions are access conscious. You have many perceptions and other mental states that you cannot verbally report. The existence of non-access-conscious mental states is one of the most surprising and well-confirmed findings of twentieth-century psychology. Access-conscious states are only the tip of the iceberg in our mental life.

A fifth thing we might mean by ‘consciousness’ is phenomenal consciousness or qualia (see Nagel 1974). This is harder to pin down, but it is central to our concept of consciousness. To understand what phenomenal consciousness is, imagine taking a god’s eye view of your mental life. There are lots of events taking place inside your head at any given moment. You have beliefs (that Paris is the capital of France) and desires (to eat lunch soon). You plan (to go to the cinema), and your plans result in motor actions (turning the handle on your front door). You perceive (this book), and you make perceptual discriminations between objects in the environment (between the book and the background). But there is something else going on. Your mental activity isn’t just information processing ‘in the dark’. It is accompanied by *subjective feelings*. Imagine that a piece of dark chocolate is placed on your tongue. Now imagine that instead a breath mint was placed on your tongue. You could, of course, tell the difference between these two stimuli. But there is more going on than mere discrimination. It *feels* a certain way to taste chocolate; it *feels* a certain way to taste mint; and those two feelings are different. These conscious feelings, which accompany many aspects of our mental life, are what is meant by *qualia* or *phenomenal consciousness*. We currently have no more precise definition than this of what ‘qualia’ means. The best we can do is gesture at examples. As Louis Armstrong said when he was asked to define jazz, ‘If you have to ask, you’ll never know’.

3 Controversies and progress

We have met five things we might mean by ‘consciousness’. This is not an exhaustive list; you may wish to think for yourself of other ways in which we use the term ‘consciousness’. The list is also a work in progress. The science of consciousness is in its infancy, and it is too early to say whether this is the correct way of splitting up our folk concept into natural kinds for scientific study. One particular area of controversy concerns phenomenal consciousness. Daniel Dennett (see the list of internet resources) has recently argued that phenomenal consciousness is not a distinct conception from access consciousness. This is a bold claim: on the face of it, being globally broadcast in the brain seems different from having qualitative feelings (one could conceive of a creature having one but not the other). Many researchers think that access consciousness and phenomenal consciousness are distinct, even if it turns out mechanisms that give rise to them are partly shared.

Let’s assume that the unpacking of the folk concept of consciousness above is correct for our purposes. For each of the different interpretations of ‘consciousness’ above, one might ask the three questions about consciousness that we posed at the beginning of this chapter – what makes us conscious, why do we have consciousness, and which creatures other than humans have consciousness?

Some of these questions turn out to be easier to answer than others. For example, we are making good progress with explanations of what it is about us, as physical beings, that makes us awake, as will we see below. However, one set of questions – those pertaining to phenomenal consciousness – have turned out to be incredibly hard to address. These questions centre on what has been called the hard problem of consciousness. Let’s take a closer look at the hard problem.

4 The hard problem

The hard problem of consciousness is to explain how it is that creatures like ourselves have phenomenal consciousness (Chalmers 1995). What is it about us, as physical beings, that produces conscious feelings? Consider yourself from two different perspectives. From a subjective point of view, you appear to know, with certainty, that you have conscious feelings. These conscious feelings come in many kinds: pains, aches, tickles, hunger pangs, itches, tingles, and tastes. Philosophers use the shorthand ‘what it is like’ to refer to the conscious feeling that tends to occur when we do a particular activity. We might talk about ‘what it is like to stub one’s toe’, ‘what it is like to eat a raw chilli’, ‘what it is like to hold a live mouse cupped in one’s hands’. In each case, what is meant is the conscious experience that usually

occurs when we are doing this activity. The ‘what it is like’ locution provides us with a way of referring to conscious feelings that we may not already have a name for. Reflecting on our conscious experience using ‘what it is like’ reveals that we already know a lot about the nature and structure of our conscious experience. What it is like to taste chocolate is *different* from what it is like to taste mint. What it is like to taste clementines is *similar* to what it was like to taste oranges. What it is like to taste lemonade is *more similar* to what it is like to taste limeade than it is to what it is like to taste coffee. Our conscious feelings have a definite structure and they bear relations to one another. Conscious experiences are not randomly distributed in our mental life. There is a lot to be discovered about our conscious experience from a subject’s point of view. The study of conscious experience from a subject’s point of view is called ‘phenomenology’.

Now, imagine viewing yourself from an outsider’s perspective. From this standpoint, it seems utterly remarkable that your brain produces conscious feelings at all. If we did not know this from our own experience, we would never have guessed it. Consider your brain as a physical object. Your brain is made up from over a hundred billion neurons wired in a complex web that drive your muscles using electrochemical activity. How does activity in this network produce a conscious experience? We might imagine, at least in rough outline, how activity in this network could store information, discriminate between stimuli, and control your behaviour. We don’t have a full story about this, but we can at least imagine how such a story might go. The case of conscious experience is different. We have no idea where to start to explain how activity in the brain produces conscious feelings. We do not even know the rough shape such an explanation may take. We know, from our phenomenology, that we have conscious experiences and that these experiences have a rich structure. But we have no idea how to explain what it is about us, as physical beings, that produces this.

It is helpful to divide the hard problem into two parts. The first part of the hard problem is to explain why physical activity in our brains produces conscious experience *at all*. Why are we not philosophical zombies – beings who have the same behaviour and information processing that we do, but for whom this goes on ‘in the dark’, unaccompanied by conscious experience? Why do we have conscious experience rather than no conscious experience at all? The second part of the hard problem is to explain, having accepted that we have conscious experience, why we have the *particular* conscious experiences that we do. Why does what it is like to taste chocolate feel to us like *this* and what it is like to taste mint feel like *that*, and not, say, the other way around? Why are our conscious feelings distributed the way they are in our mental life? Both parts of the hard problem are stunningly hard. We don’t have anything even approaching an answer to either question. Scientific work on consciousness tends to set the hard problem aside in order to concentrate on

more tractable questions.

Why is the hard problem so hard? One diagnosis is that there is an explanatory gap between the two perspectives described above: our first-person knowledge from phenomenology and our third-person knowledge from the natural sciences. Both appear to be legitimate sources of knowledge about our mental life. The difficulty is that it is hard to see how to link these two sources of knowledge together. Science has an impressive track record in unifying diverse fields of knowledge. A common pattern in science is to unify fields by reductive explanation: by explaining the phenomena of one field in terms of those of another. The kinetic theory of gases, for example, allows us to explain a wide range of phenomena concerning temperature and pressure in terms of the physical laws and mechanisms governing the constituent molecules of a gas. However, past successes at reductive explanation in science have exclusively concerned knowledge gained from the third-person point of view. Pressure, temperature, and the physical characteristics of molecules are all studied from a third-person point of view; we do not gain knowledge of them via introspection. The puzzle here is how to explain first-person conscious experience in terms of third-person studies of physical activity in the brain. There is no precedent for doing this. The task has a qualitatively different character from past reductions in science. A number of philosophers, including Frank Jackson, David Chalmers, and Thomas Nagel have argued that the particular challenges posed by this reduction mean that science will never explain conscious experience in terms of brain activity, and so the hard problem will never be solved. Let's look at Frank Jackson's argument for this claim (see Ludlow et al. 2004 for replies).

5 Thought experiment: Mary the colour scientist

Imagine that a brilliant neuroscientist, Mary, is born and grows up inside a black and white room. Mary has never seen colour, but she is fascinated by the human brain processes that detect and process colour. Inside her room, Mary is provided with encyclopedias that describe everything about the workings of the human brain. These books cover not only current knowledge in neuroscience, but all that could possibly be discovered. From her books, Mary learns every neuroscientific fact about colour vision. She understands in exquisite detail how the human brain responds to colours, and how colour information is processed by the brain.

One day Mary is released from the room. When she goes outside she spots a red rose and experiences colour for the first time. Jackson claims that at this moment Mary learns something new about human vision. Mary learns *what it is like to see red*. She already knows how the brain processes colour, but now she learns what it *feels*

like to see red: the character of the conscious experience that accompanies seeing red. Jackson puts it to us that this distinctive conscious feeling is not something that Mary could have predicted from her books. No matter how much she learned about human visual processes inside her room, she would not have known what it is like to see red until she had the experience herself. When she leaves the room, Mary learns, via her first-person introspection, a new fact about the human visual system. This fact was not contained within, or deducible from, her neuroscientific knowledge. Now, Jackson says, if Mary could not have predicted in advance what the experience of seeing red would be like, then we will never be able to explain conscious experience in terms of brain activity. Even if we were lucky enough to have a complete neuroscience, we would at best be in Mary's predicament. We would not be able to show how the physical facts give rise to phenomenal consciousness. No matter how far our neuroscience progresses, we will not be able to explain how brain activity gives rise to conscious experience.

At the moment, science is unable to address the fundamental question of how physical activity yields phenomenal consciousness. A great deal of progress, however, has been made in resolving questions about the sorts of brain activity and psychological functions that are *correlated* with phenomenal consciousness. In time, this growing body of knowledge may contribute to an understanding of the issues raised by the hard problem. In the remainder of this chapter we will review recent findings in psychology and neuroscience that, over the last few decades, have advanced our understanding of these profound questions. We will describe the questions that consciousness science currently finds relevant, and go on to discuss how states of consciousness arise from brain activity, and what determines the content of consciousness – our awareness of ourselves and our environment – at any given time.

6 Scientific perspectives: The questions consciousness scientists ask

So what sorts of questions do scientists ask when they investigate consciousness? And how much progress have researchers made in turning these questions from general musings into enquiries that can be investigated empirically? The most fundamental question is how the activity of a physical system – the brain and central nervous system – can create consciousness and subjective experience (qualia) in the first place. As we saw above, this is one of the formulations of the hard problem, and science is no closer than it has ever been to answering it. The reason is that we have no idea what an answer would look like; it may be staring us in the face, and we simply don't have the tools – the conceptual framework – to recognize it.

Other major questions are considered ‘easy problems’, but it’s important to be clear about what ‘easy’ means in this context. As psychologist Steven Pinker (2007; see the list of internet resources) aptly put it, these questions are only easy in the same way that curing cancer is easy – it’s not that they are easy to solve, but rather that it would be easy to recognize an answer when we found one. The two main ‘easy’ questions that scientists are currently interested in both come under the heading of neural correlates of consciousness (abbreviated as NCC). Both questions are concerned with figuring out what sorts of brain activity correlate with – i.e. happen at the same time as – processes related to consciousness. The first of these asks what neural activity determines an individual’s *state* or *level* of consciousness – what patterns of brain activity lead to a person being awake or asleep, in a coma or vegetative state, and so on. Researchers who investigate this question are also interested in what each of these states means – what level of information processing can occur in the brain when a person is in each state, and to what extent patients with disorders of consciousness, such as those in a vegetative state, can still have some capacity for subjective experience.

The second question asks what processes shape the *content* of our consciousness – our momentary awareness of ourselves and of the world around us – at any given point in time? A lot of recent consciousness research has focused on perceptual awareness – discovering the link between what the world presents to us through our senses and what we become aware of (you can think of what we become aware of as related to the qualia we mentioned above – but science is currently more interested in what happens in the brain when qualia occur, rather than with how the brain can create qualia at all).

7 States of consciousness

7.1 What different states of consciousness are there?

How does brain activity give rise to different states of consciousness? Let’s start by examining what states exist. It’s useful to think about a person’s state of consciousness not as single thing that there can be either less or more of, but rather as a combination of two separate factors: wakefulness (*level* of consciousness) and awareness (having conscious *content*; see Figure 7.1).

Our level of wakefulness determines whether we are awake or not. Our awareness is our capacity to think, feel, and perceive ourselves and the environment around us. This awareness is what allows us to interact with the external world in a meaningful way. It may seem strange to separate wakefulness and awareness, but as we will see,

it enables us to pinpoint the differences between various states of consciousness (Laureys 2007).

Figure 7.1 States of consciousness. The different states are organized by both wakefulness and awareness. REM, rapid eye movement. (Image taken from S. Laureys, *Scientific American* May 2007. Copyright: Melissa Thomas.)

Right now your state of consciousness is at the high ends of both the wakefulness and awareness scales; as you fall asleep tonight, you will first become drowsy, and eventually fall into a deep sleep, a state in which both your wakefulness and awareness will be low. Under artificially induced anaesthesia, or trauma-induced coma, wakefulness and awareness are reduced even further. If these were the only states that existed, there would be no need for two separate axes to describe them. There are, however, states in which one of the factors is high while the other is low. The most obvious example is dreaming, where wakefulness is low (a person is asleep) but awareness is high (the person experiences thoughts, feelings, and sensations). In the rare state of lucid dreaming, people are not only asleep and dreaming, but become aware of the fact that they are dreaming. Unfortunately, there are also clinical states known collectively as disorders of consciousness, where brain injury leads to reduced awareness during wakefulness. These disorders include the vegetative state and the minimally conscious state. Patients in a vegetative state have normal sleep-wake cycles, but when they are awake (with their eyes open) they don't respond to their environment and don't produce any meaningful behaviour (such as following instructions, communicating, or moving in a way that would indicate they know what's going on around them). As far as anyone can tell, they seem to be experiencing no thoughts or feelings. Occasionally, vegetative state patients' conditions improve and they are reclassified as being in a minimally conscious state, a condition in which they occasionally exhibit limited responsiveness to their environment.

Finally, it is worth mentioning the tragic condition known as locked-in syndrome. This is not a disorder of consciousness, and locked-in patients are fully awake and aware; their brain injury has rendered them unable to control their body, so they can't move or communicate. Sometimes, such patients retain limited control of a small number of muscles, and can use them to communicate. Some experts estimate that as many as 40 per cent of vegetative state diagnoses may be mistaken, with the patients in fact retaining some awareness but not the ability to communicate (Monti et al. 2010).

7.2 States of consciousness and the brain

What sorts of brain activity determine a person's state of consciousness? As far as we know, there is no specific brain area whose activity is solely responsible for either wakefulness or awareness. The brain is a vastly integrated system, and its state is the outcome of many subsystems' combined activity. There are, however, brain areas that are known to contribute to specific aspects of consciousness. Wakefulness is highly dependent on activity in subcortical regions (regions that lie deep in the brain, beneath the outside layer, which is called the cortex). These subcortical areas include the reticular formation, which is located in the midbrain, an area at the bottom of the brain just above the spinal cord. The midbrain is part of the brainstem, an evolutionarily ancient neural structure (meaning it is similar in humans and in many other animals, indicating it evolved long ago in our mutual ancestors). The reticular formation is part of a network of areas known as the reticular activating system, which regulates sleep–wake cycles. Another subcortical region involved in regulating wakefulness is the thalamus, which serves as a general relay station for information transmitted in the brain, and is important in regulating arousal (how alert we are during wakefulness). Damage to any part of the reticular activating system or certain parts of the thalamus can lead to coma or disorders of consciousness such as vegetative state, but these disorders can also result from damage to many other brain regions.

Unlike wakefulness, which depends on subcortical structures, the presence of awareness is largely related to cortical function (the cortex is a relatively recent evolutionary development, and is responsible for higher mental functions in humans). Awareness can be thought of as consisting of two complementary elements. The first is awareness of the external environment; this is the awareness we have whenever we need to navigate through our environment, interact with other people, or do anything else that requires the use of perceptual information. Brain regions involved in this type of awareness comprise a network in the frontal and parietal lobes of the cortex, mostly located in the upper-outer parts of the brain's surface. These areas are known collectively as the task-positive network, fronto-parietal network, or dorsal attention network.

The second element of awareness is the kind that occurs when we are not focused on the external environment but on our own inner world – daydreaming, retrieving memories, or planning for the future. Doing these things involves activity in a network of regions known collectively as the task-negative network or default-mode network (because it becomes active when we're not performing any specific task related to the outside world). This network comprises cortical regions that are mostly located on the medial surface of the brain (the inside part, where the brain's

two hemispheres face each other).

When we are awake, we are usually either focusing on something in the external environment or directing our attention inward; it's rare (and some would say, impossible) to be doing both things at the same time. It is not surprising, therefore, that the activity of the task-positive and task-negative systems is negatively correlated – when either is high, the other is low. Although changes in wakefulness are mostly governed by subcortical structures, as we saw above, such changes affect cortical activity, too. As we descend from full wakefulness to sleep, activity throughout the cortex changes. During wakefulness, different parts of the cortex are busy communicating with each other. This is known as functional connectivity, and can be seen in measures of coordinated activity between brain regions. As we fall asleep, this communication is sharply reduced. The greatest reduction occurs in the coordination between frontal and posterior (back) regions of the cortex (note that these are changes in functional, not structural connectivity – in other words, the physical connections between different parts of the cortex remain intact, but those different parts don't communicate with each other as much).

7.3 Awareness in disorders of consciousness?

One day in 2005, a young woman was injured in a car accident. She sustained severe brain damage, and was in a coma for a while. She awoke from her coma, but did not regain consciousness. She was diagnosed as being in a vegetative state. Despite having sleep-wake cycles, and spending her days with her eyes open, she didn't respond to any attempt at communication, and made no purposeful movements. As far as anyone could tell, she displayed no awareness whatsoever.

In early 2006, a group of researchers from the UK and Belgium performed a functional MRI scan of this patient. During the scan, they gave her instructions. Some of the time, they asked her to imagine she was playing tennis; at other points, they asked her to imagine walking through the rooms of her house. They had tried this before with other vegetative state patients, but had gotten no response. This time was different, though. The woman's brain activity changed markedly depending on the instructions she got. When asked to imagine playing tennis, her supplementary motor area, a cortical region involved in planning movements, became active. When asked to imagine walking through her house, the activated regions included areas known to be involved in spatial navigation, such as the parahippocampal place area. Most importantly, her neural responses were indistinguishable from those observed in a group of healthy control participants. The researchers concluded that despite the absence of observable signs of awareness, the woman's responses to instructions meant that she possessed a certain level of awareness. They published the study

under the provocative title ‘Detecting Awareness in the Vegetative State’ (Owen et al. 2006).

The paper created quite a stir. Was the woman indeed conscious? Could vegetative state patients, in general, be conscious? Critics were unconvinced. The pattern of brain activity seen in this patient, they said, does not necessarily indicate awareness. The observed activity might simply be an indication of how much the brain can do without awareness – perhaps an automatic response to hearing the words ‘tennis’ and ‘house’ – rather than evidence that this patient is aware. A related criticism addressed the logic underlying the researchers’ conclusions: just because all cucumbers are green, this doesn’t mean that anytime you see something green it must be a cucumber; likewise, if imagining walking through one’s house leads to a certain pattern of brain activity (in healthy people), this does not mean that anytime you see this pattern of activity it means a person (our patient) is imagining that same thing.

The logic of these criticisms is sound, so the researchers went on to provide a stronger case. In a new study, the same imagery tasks were used – but this time, they were used for communication (Monti et al. 2010). Each task was associated with a specific answer. Vegetative state patients were asked simple yes/no questions (to which the researchers knew the answers – for example, ‘do you have any sisters?’), and had to think of playing tennis if the answer was ‘yes’, and of walking through their house if it was ‘no’. The researchers found one patient (not the one from the first study) whose neural responses provided the correct responses to questions despite the absence of any behavioural evidence of awareness. The researchers concluded that this is reliable evidence of awareness, as the connection between the answer (yes/no) and the actual response (tennis/house imagery) was arbitrary, making it very hard to believe automatic activation could be at play.

This seems sensible, but it is important to remember that the findings do not indicate that all vegetative patients are conscious – out of 54 patients in the follow-up study, only 5 showed differential activity related to the mental imagery tasks, and only 1 was able to use these tasks to communicate. Nonetheless, it seems there is more going on than we previously realized, at least in some cases, under vegetative patients’ unresponsive exterior.

7.4 Perceptual awareness

So far we’ve focused on states of consciousness and the brain activity that underlies them. But as we mentioned earlier, consciousness researchers are also interested in the processes that determine the specific content of our consciousness at any given

time – our perceptual awareness.

Let's start with our subjective awareness of the environment. Despite normally thinking that we are aware of many different things, research suggests that at any given moment we are only aware of a small subset of the information entering our brain through the senses. There are some great demonstrations that attest to the limitations of our awareness – if you want to experience this yourself, see the internet resources at the end of the chapter for directions to videos made by Daniel Simons and Christopher Chabris (see also their papers from 1999 and 2010). The phenomenon demonstrated in these videos is called inattention blindness (Mack and Rock 1998), and its very existence reveals the intimate connection between awareness and attention. A related, but slightly different phenomenon, known as change blindness occurs when no new elements enter or leave the display, but a change in an existing element goes unnoticed.

Hundreds of studies have looked into both inattention and change blindness, in an effort to figure out what determines the things we are likely to miss. One of the relevant factors seems to be our 'attentional set' – if we're not looking for something, there's a greater chance we won't notice it (for example, have you ever encountered your next-door neighbour in a different part of town, and walked right past them without saying hello just because you hadn't expected to see them there?). Another relevant factor seems to be the capacity limits of our visual working memory (the store of visual information that is available for our immediate use). In studies of change blindness, researchers often present colourful shapes, remove them briefly and return them, and see whether observers are able to notice when one of the shapes has changed. By presenting displays with different numbers of shapes, researchers can find out how much they can increase the number of shapes before people start making mistakes – and the number isn't large at all; it depends on the type of object (and on the specific change), but in most studies it's around four.

7.5 Selecting information for awareness

What are the neural processes that choose which information to bring into awareness? When investigating such questions, researchers have made extensive use of a type of stimuli known as bistable images. The most well-known example of a bistable image is the Necker cube (see Figure 7.2A), which has two possible visual interpretations, each with a different side seen as being at the front.

The Necker cube has the three hallmarks of bistable images. First, it is a single image that is associated with two possible conscious interpretations; second, these two

interpretations cannot be seen at the same time (try it!); and finally, the two interpretations tend to alternate. Another example of a bistable image is the Rubin face–vase (Figure 7.2B). Why are bistable images so useful to consciousness researchers? They provide a great opportunity to isolate awareness from the external, physical stimulus: the image itself doesn't change, but our conscious interpretation of it does. As the only change is happening in our brain, figuring out the neural mechanism for this change would provide insight into how the brain selects perceptual inputs for conscious representation.

Figure 7.2 Bistable images. (A) Necker cube. (Copyright image: drawn by Carmel) (B) Rubin face–vase.

Several kinds of bistable images have been used in neuroimaging studies, where researchers examined what changes in the brain are time-locked to observers' reported changes in perception. Repeatedly, researchers have found that during perceptual switches, activation can be seen not only in the visual cortex (which is located in the occipital lobe, at the back of the head), but also in certain parts of the frontal and parietal lobes already mentioned earlier in this chapter (Rees et al. 2002). So can we conclude that the fronto-parietal network is responsible for choosing which images enter awareness?

Not so fast. Just because something happens in the brain at the same time as a reported perceptual event (such as a switch in a bistable image), this doesn't mean the brain activity is *causing* the change – it correlates with the change in perception, but correlation is not causation. Let's say we observe activity in parietal cortex at the time of a switch. This could, for example, indicate that rather than triggering the switch, parietal cortex is involved in noticing that a change is happening; or it could be that something else – say, activity in a different region like occipital cortex – caused both the perceptual change *and* the activity in parietal cortex. In order to go beyond correlational evidence, it is necessary to manipulate the factor you suspect might be having a causal effect (in this case activity in certain brain regions), and see how the manipulation affects the thing you're interested in (in this case, perceptual switches in bistable images). In recent years, several researchers have used a technique called transcranial magnetic stimulation (TMS) for this purpose. TMS works by applying a powerful magnetic pulse to the surface of the head; for a very short time, this interferes with the activity of the area of cortex directly under the pulse. Interestingly, a series of studies has revealed that applying TMS to adjacent areas within parietal cortex can lead to completely different results – in some cases making the rate of perceptual switches faster, and in others slowing it down (Carmel et al. 2010; Kanai et al. 2011). This provides strong evidence that parietal cortex is causally involved in bistable switches. But as is almost always the case in research, the emerging picture is more complex than we'd expected: the next step would be

to figure out the specific roles of the different parts of cortex whose stimulation leads to different effects, and how the neural system as a whole comes up with a consensus that is represented in awareness.

7.6 Suppression from awareness

To understand consciousness, we need to know the differences between processes that require awareness and those that don't. If we can perceive something without awareness, this tells us awareness is not *necessary* for such perception, and narrows down the list of processes awareness *is* necessary for. So how do we investigate unconscious perception? Researchers have developed several techniques that allow them to present stimuli that enter the relevant sense organs, but which the observer does not have conscious access to and is unable to report. One widely used example is visual masking, in which a visual image is presented very briefly and followed immediately by another image – the 'mask' – that is presented for longer. People are often unable to report the first image, and may even deny there was one at all. One of the original studies that used masking to demonstrate unconscious perception employed a method called 'masked priming': observers saw words that were followed immediately by a mask (a meaningless pattern), which prevented awareness of the words. After each masked word, observers were shown a string of letters and had to decide if it was a real word or not. Interestingly, people spotted a real word faster when it was semantically related to the masked word (for example, the word 'child' following a masked 'infant'), than when it was not (e.g. 'orange' following 'infant'). This indicated that masked words were processed deeply enough to activate a semantic network in the brain (enabling faster recognition of related words), despite remaining unavailable to awareness; this 'priming' effect was just as large without awareness as with it (Marcel 1983).

A different method, called continuous flash suppression (CFS), allows for displays lasting several seconds while ensuring observers do not become aware of what they see. In CFS, a strong – high-contrast and rapidly changing – image is shown to one eye, designated the dominant eye, while the other, suppressed eye views a weaker stimulus (which has lower contrast, but would still be visible if viewed on its own). Under suitable conditions, the weaker image will not enter awareness, despite the suppressed eye being continuously exposed to it (Tsuchiya and Koch 2005). A recent study used CFS to investigate classical fear conditioning with and without awareness (Raio et al. 2012). Observers were shown two different pictures, for four seconds each, several times in random order. At the end of the four seconds, one of the pictures – always the same one – was occasionally paired with a mild (but unpleasant) electric shock to the wrist. The participants' skin conductance

response (a physiological measure, basically indicating how much you sweat) was measured to examine the development of the characteristic fear response – higher skin conductance whenever the image that predicts a shock is shown. The study included two separate groups: in one, participants were aware of the images; for the other, the images were suppressed from awareness by CFS. Both groups developed a fear response. Interestingly, however, this response developed differently over time: The unaware group's fear response actually arose faster than the aware group's, but the fear learning didn't 'stick' – by the end of the experiment, the greater response to the threatening image had disappeared. For the aware group, learning developed more gradually, but was stable. This qualitative difference – different time courses for conscious and unconscious fear learning – may tell us something fundamental about the role of awareness: we may not need it to form an association between a stimulus and a threat, but for this association to become stable, further processes – ones that involve awareness – may be required.

8 Theories of consciousness

We still have no idea why neural activity should be accompanied by consciousness at all. However, as we have now seen, science has made quite a bit of progress in characterizing the neural activity and cognitive functions that are associated with conscious experience. This progress has led to a number of theoretical ideas on the kinds of neural processes that lead to subjective awareness. In this section we will briefly describe three of the theories that are currently most prominent.

Global neuronal workspace theory (Dehaene et al. 2003) proposes that in order for perceptual inputs to reach awareness, two conditions must be met. First, the activation that the external stimulus causes in 'early' regions of the brain (those devoted to perceptual processing from a particular sense, for example vision) must be strong enough. Second, the perceptual information must be shared with other 'modules' (neural systems devoted to other kinds of processing) across the brain. According to this theory, attention is the crucial process that takes perceptual processing and transmits it to the 'global workspace', where it becomes available to other systems.

A different theory attributes conscious experience to recurrent processing (Lamme 2010). This theory focuses on the dynamic flow of perceptual information in the brain. The first stage in perception is a 'feedforward sweep', where the sensory information makes its way up a hierarchy of brain areas devoted to analysing it – for example, visual information is conveyed from the eyes to primary visual cortex, where it undergoes basic analysis; it then goes on to secondary visual cortex for

further detailed analysis, and then on to areas that specialize in specific aspects of vision such as motion or colour. This entire feedforward sweep is not accompanied by awareness. However, as it progresses, there is another process that takes place in parallel: feedback loops become active, so that each area that receives input also communicates with the region that sent the information, adjusting and fine-tuning the activity of the previous area to improve the quality of data it sends, resolve ambiguities, or settle contradictions. This feedback is called recurrent processing, and each level at which it occurs contributes to awareness.

The third prominent theory is information integration theory (Oizumi et al. 2014). Unlike the two theories described so far, which focus on perceptual awareness, this theory attempts to quantify the relation between consciousness in general and the way information is stored in the brain (or in any other system). According to the theory, consciousness is a continuum – there can be more or less of it – and the amount, or level, of consciousness in a given system is determined by how much information it integrates. What does it mean to integrate information? Well, information is integrated when you can't get all of it just by looking at the individual parts of the system. You can have lots of information that is not integrated: a digital camera, for example, can record colour values for millions of pixels, but it is not conscious; according to the theory, this is because there is no integration of all this information – none of those pixels are connected to each other, and no information is passed between them. However, sharing doesn't automatically entail integration: for example, it wouldn't be a good idea for every neuron in the brain to be connected to every other neuron, because then any activity would cause chaos, with all neurons becoming active. The systems that store the most information are those that both integrate information by sharing it, and differentiate this information, making the system's state unique amongst all the possible states it could be in. According to the theory, integrated information and consciousness are *the same thing*. It's important to understand that this is an *assumption* of the theory, not an outcome of its calculations. Does consciousness really come down to nothing more than information organized in a particular way? Perhaps; the theory, however, doesn't lead to this conclusion, but rather uses it as its starting point.

At this point there is no theory that offers a full, unified account of consciousness and how it arises from the activity of physical systems. Current theories offer agendas for future research – what ideas and issues we should be pursuing if we want to understand consciousness. Time will tell whether these directions will turn out to be fruitful, or whether future developments will suggest other directions.

9 Chapter summary

- The hard problem of consciousness is to explain how our brains produce phenomenal consciousness. We know that we have phenomenal consciousness from our own subjective experience, but we have no idea how brains produce it.
- A number of philosophers, including Frank Jackson, have argued that science will never solve the hard problem of consciousness. Scientific research on consciousness currently lacks the conceptual framework needed to address the hard problem. It therefore focuses on 'easy' problems (that are only easy in the sense that we would recognize an answer when we found one).
- The 'easy' problems of consciousness that scientists are most interested in concern the neural correlates of consciousness: how does brain activity determine *states* of consciousness, and what neural and psychological processes determine the *content* of consciousness at any given time?
- States of consciousness can be categorized as a combination of two factors: wakefulness and awareness. When we are awake we have a high level of both factors; when we are asleep we are low on both; during dreaming, we have a high level of awareness, but a low level of wakefulness; and vegetative state patients have a high level of wakefulness with a low level of awareness.
- There is evidence from brain imaging that some patients in a vegetative state may retain some awareness of their environment, despite an absence of any behavioural indications.
- Perceptual awareness is the term given to the sensory information we are aware of at any given moment. Despite our subjective sense that we have a rich, detailed awareness, phenomena such as inattentional blindness and change blindness demonstrate that at any given time, our awareness is very limited.
- Bistable images are an important tool in studying awareness, because our perception of them can change without any change in the images themselves.
- A great deal of perception and cognitive processing can occur without awareness. Studying unconscious perception is thus an important way of distinguishing processes that require awareness from those that don't, and of finding out which processes may differ in the way they are carried out with and without awareness.
- Several theories have been proposed to explain various aspects of conscious-

ness, though none of them currently offers a full account: the *global neuronal workspace model* and *recurrent processing theory* both offer accounts of the way perceptual awareness arises, whereas *integrated information theory* suggests that consciousness can be measured as the amount of integrated information in a system.

Study questions

1. What is the difference between folk concepts and scientific concepts that pick out natural kinds?
2. What are the various things that folk may mean by 'consciousness'? Can you give a simple example for each?
3. What is the hard problem of consciousness, and why is it hard?
4. What is Jackson's argument that science will never solve the hard problem of consciousness? Do you see any flaws in Jackson's argument?
5. How are different states of consciousness defined, and what determines a person's current state?
6. What is the evidence for awareness in the vegetative state? Is it convincing? Is it possible to know with certainty whether a vegetative patient is aware, and if so, what evidence would such certainty require?
7. What is the relation between perceptual awareness and other mental faculties, such as attention and memory?
8. Why are bistable images useful tools in research on perceptual awareness?
9. What can evidence of unconscious perceptual processing teach us about consciousness?
10. What aspects of consciousness do current theories propose explanations for? And what type of problems – hard or easy – do they address?

Introductory further reading

- Chalmers, D. J. (1995) 'Facing up to the problem of consciousness', *Journal of Consciousness Studies* 2: 200–19. (A great description of the hard problem of consciousness.)

- Laureys, S. (2007) 'Eyes open, brain shut', *Scientific American*, May, pp. 32–7. (An engaging and accessible review of what is known about states of consciousness and the brain.)
- Ludlow, P., Nagasawa, Y. and Stoljar, D. (eds) (2004) *There's Something about Mary*, Cambridge, MA: MIT Press. (This collection includes Frank Jackson's original paper with his Mary argument and many excellent responses. Highlights include the responses to Jackson by David Lewis and Daniel Dennett.)
- Mack, A. and Rock, I. (1998) *Inattentional Blindness*, Cambridge, MA: MIT Press. (An influential book that introduced the phenomenon of inattentional blindness.) 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: 1402. (This simple, one-page paper reported the vegetative patient whose brain activity indicated she may be aware of her surroundings.)
- Rees, G., Kreiman, G. and Koch, C. (2002) 'Neural correlates of consciousness in humans', *Nature Reviews Neuroscience* 3: 261–70. (This review paper is no longer new, but covers the logic and fundamental findings of research on perceptual awareness in an accessible and engaging way.)

Advanced further reading

- Block, N. (1995) 'On a confusion about a function of consciousness', *Behavioral and Brain Sciences* 18: 227–47. (Block's original paper in which he draws the access/ phenomenal consciousness distinction.)
- Carmel, D., Walsh, V., Lavie, N. and Rees, G. (2010) 'Right parietal TMS shortens dominance durations in binocular rivalry', *Current Biology* 20: R799–800. (This study demonstrates that different regions within parietal cortex play different roles in selecting how visual information will be represented in awareness.)
- Dehaene, S., Sergent, C. and Changeux, J.-P. (2003) 'A neuronal network model linking subjective reports and objective physiological data during conscious perception', *Proceedings of the National Academy of Sciences* 100: 8520–5. (This widely cited paper introduces the global neuronal workspace model.)
- Kanai, R., Carmel, D., Bahrami, B. and Rees, G. (2011) 'Structural and functional fractionation of right superior parietal cortex in bistable perception', *Current Biology* 21: R106–107.
- Lamme, V. A. F. (2010) 'How neuroscience will change our view on consciousness', *Cognitive Neuroscience* 1: 204–20. (A good introduction to recurrent processing theory, and a detailed description of the challenges facing the

cognitive neuroscience of consciousness.)

- Marcel, A. J. (1983) 'Conscious and unconscious perception: experiments on visual masking and word recognition', *Cognitive Psychology* 15: 197–237.
- Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D. et al. (2010) 'Willful modulation of brain activity in disorders of consciousness', *New England Journal of Medicine* 362: 579–89. (In this study, researchers found a vegetative patient who was able to use his brain activity to answer questions.)
- Nagel, T (1974) 'What is it like to be a bat?', *Philosophical Review* 83: 435–50. (A wonderfully prescient paper that makes vivid what later came to be known as the hard problem of consciousness.)
- Oizumi, M., Albantakis, L. and Tononi, G. (2014) 'From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0', *PLoS Computational Biology* 10: e1003588. (The most up-to-date version of integrated information theory.)
- Raio, C. M., Carmel, D., Carrasco, M. and Phelps, E. A. (2012) 'Unconscious fear is quickly acquired but swiftly forgotten', *Current Biology* 22: R477–9. (This study used CFS (continuous flash suppression) to investigate whether a new fear can be acquired without awareness, and showed that it can – but that the learning has a different time course than conscious learning.)
- Simons, D. J. (2010) 'Monkeying around with the gorillas in our midst: familiarity with an inattentional-blindness task does not improve the detection of unexpected events', *i-Perception* 1: 3–6.
- Simons, D. J. and Chabris, C. F. (1999) 'Gorillas in our midst: sustained inattention blindness for dynamic events', *Perception* 28: 1059–74. (This study, as well as Simons 2010, use insightful and entertaining demonstrations of inattentional blindness and change blindness.)
- Tsuchiya, N. and Koch, C. (2005) 'Continuous flash suppression reduces negative afterimages', *Nature Neuroscience* 8: 1096–1101.

Internet resources

- David Chalmers explaining the hard problem of consciousness: conscious-pictures (2010) 'David Chalmers on the “hard problem” of consciousness – Chronicles of Consciousness', 8 April, YouTube [video-streaming site], <http://www.youtube.com/watch?v=kdbS-HUAXC8>
- Daniel Dennett's recent lecture in which he criticizes the access/phenomenal consciousness distinction: Dennett, D. (2013) 'Consciousness about access and consciousness', *Consciousness Online*, 15 February [podcast, conference

website] <http://consciousnessonline.com/2013/02/15/on-a-phenomenal-confusion-about-access-and-consciousness/>

- Great overview of philosophical problems concerning consciousness with many suggestions for further readings: van Gulick, R. (2014) 'Consciousness', in E. N. Zalta (ed.) Stanford Encyclopedia of Philosophy (Spring 2014 edn) [online encyclopedia], <http://plato.stanford.edu/entries/consciousness/>
- The hard problem versus the easy problems of consciousness: Pinker, S. (2007) 'The brain: the mystery of consciousness', Time, 29 January [magazine with online access], <http://content.time.com/time/magazine/article/0,9171,1580394,00.html>
- Limitations of perceptual awareness: Simons, D. (2010) 'The monkey business illusion', YouTube, 28 April [video-streaming site], http://www.youtube.com/watch?v=IGQmdoK_ZfY&list=PLB228A1652CD49370&index=2; Simons, D. and Chabris, C. F. (uploaded by Simons, D. 2010) 'Selective attention test', YouTube, 10 March [video-streaming site], <http://www.youtube.com/watch?v=vJG698U2Mvo&list=PLB228A1652CD49370>