The philosophy of phenomenal consciousness

An introduction

Zoe Drayson
University of Stirling, Stirling

The scientific study of consciousness is constantly making new discoveries, but one particular aspect of consciousness remains problematic to explain. This is the fact that conscious experiences present themselves to us in a first-person way: there is something it feels like to be the subject of a conscious experience. This 'phenomenal' aspect of consciousness seems to be subjective, private, and knowable in a special way, making it difficult to reconcile with the scientific focus on objective, third-person data. This introduction provides an overview of phenomenal consciousness, explores philosophical arguments about its nature, and considers whether or not we should expect to find an explanation for the properties of phenomenal consciousness.

1. Introduction

Human consciousness remains largely mysterious to scientists and philosophers alike. Part of the problem doubtless comes down to technological and experimental constraints, and we can be confident that advances in brain-imaging and new experimental techniques will continue to yield more data. But another part of the problem seems to lie in the phenomenon to be explained, and how we understand it. For instance, the term 'consciousness' is used in a multitude of ways. First, creatures can be considered conscious or not conscious: we generally assume that humans are conscious creatures while bacteria are not. But conscious creatures can be in unconscious states, as when asleep or in a coma, for example. And even when a conscious creature is in a conscious state, they might not be *conscious of* a particular stimulus: 'being conscious' can be used both transitively and intransitively. These are, of course, terminological issues, and easily settled by clarifying our terms. But even once the terminology is settled, providing explanations of consciousness seems to face challenges that do not arise for other scientific problems. Conscious experiences have *phenomenal* properties: they present themselves to us in a first-person way, quite unlike the third-person data that are familiar to science.

DOI 10.1075/aicr.92.11dra

In this introduction to phenomenal consciousness, I'll begin by distinguishing the phenomenal aspect of consciousness from related notions, and showing how scientific advances have left it unexplained. In Section 3, I'll analyse the notion of phenomenal consciousness in more detail, considering its subjectivity, its privacy, the special knowledge we seem to have of our phenomenally conscious states. Some philosophers think that any states possessing such properties can't be physical states, and therefore can't be scientifically explained. In Section 4, I introduce the physicalist and non-physicalist views of consciousness, looking at the varieties of physicalism and the two main arguments for non-physicalism. Section 5 explores philosophical approaches to theories of consciousness that attempt to explain what consciousness is and how it arises. In Section 6, I focus on the very idea of providing explanations of consciousness, and on different responses to the claim that there is an 'explanatory gap' between our best explanations of consciousness and what we're actually trying to explain.

2. The conscious brain

In order to understand the ways in which consciousness remains problematic for philosophers, it helps to first consider the ways in which consciousness has become *less* mysterious in recent decades. Not long ago, our only knowledge of the human brain came from experiments on animal brains or from dissected human brains at autopsies; now, we can harmlessly scan the brains of living human subjects to get information about both the structure of the brain and its functioning. These advances have been accompanied by more precise and ingenious ways of measuring non-neural behavior (such as eye movements, reaction times, and verbal reports), cleverly-designed experiments, and enhanced data-analysis techniques. As a result, there are new discoveries being made about our mental mechanisms on a daily basis.

Consider, for example, how the eyes and brain build up our conscious visual experience of the world. When we visually experience the world, we can overtly focus our attention on different aspects of the visual scene. But our experience is also shaped by very fast 'saccadic' eye movements between foveal fixation points. (See Liversedge & Findlay, 2000, for more on saccadic eye movements and their relation to cognition.) Cutting-edge technology in the form of eye-tracking equipment can measure the location and duration of fixation, and monitor the direction of the saccades. Using these measurements of attention-switching and scan-paths, psychologists can demonstrate how our visual system builds up the information made available to us in conscious experience. (For a more detailed account of visual system anatomy and physiology, see chapter by Price, 2013, in the companion volume.)

Scientific developments have also been made in the study of pain, a paradigm state of conscious experience. We now understand the role of neurotransmitters, the chemicals that relay messages across the brain's synapses, in the pain of migraine headaches. During a migraine, levels of the neurotransmitter serotonin drop significantly, which causes blood vessels to dilate, resulting in extreme pain. (Data on serotonin and

migraine are reviewed in Panconesi, 2008.) With this knowledge of how the pain is produced, doctors can intervene to change it: migraine sufferers can be treated with drugs that optimize serotonin levels in the brain.

In addition to conscious sensations, human beings possess consciousness of themselves in the form of self-awareness. While we don't expect other animals to share our extensive capacity for reflective self-aware thought, behavioral psychologists have devised an experiment to show that some creatures have a basic form of self-consciousness. Experimenters mark an animal with an odorless dye on the front of their body, and then observe the creature in front of a mirror. Some animals, such as chimpanzees, poke at the marking or move to get a better view of the marking in the mirror, suggesting that they recognize the reflection as themselves. (Some of the most recent mirror-test experiments have been done on rhesus monkeys: see Rajala, Reininger, Lancaster, & Populin, 2010.)

Given these scientific advances in our understanding of conscious states, it is tempting to think that we're well on the way to a complete science of consciousness. But there is an aspect of consciousness that none of the above results touch upon: the way that our conscious states appear to us 'from the inside'. Work on eye movements, for example, tells us how our visual experience of a scene is built up by the way we attend to it; but none of this tells us what it feels like to be looking at an optical illusion and suddenly experience a change – for example, when looking at the duck-rabbit ambiguous image and switching from seeing it as a duck to seeing it as a rabbit, or when switching between two binocularly rivaling images. And work on the role of neurotransmitters in migraine pain can enable interventions to ease or even prevent the pain, but it doesn't tell us what it's like to experience a migraine headache, or how that feeling differs from other kinds of pains associated with different ailments. Finally, experiments like the mirror test help us to understand the cognitive abilities of non-human animals, but they don't seem to explain why, when I look in the mirror and recognize myself, my realization that the reflection is *me* is accompanied by a sensation of recognition.

What's going on here is that our conscious experiences tend to have two sets of properties: a set of causal or functional properties, and a set of 'phenomenal' properties. The causal or functional properties of conscious states are exhibited in the interactions into which they enter. Notice, for example, that conscious states like perception, pain, and self-awareness seem able to guide our behavior (including our speech and our thought) in a way that unconscious states don't. We can verbally report our pains, for example, we can use our perceptions of the world to update our beliefs, and we can reflect on our self-awareness. Our non-conscious states, on the other hand, play isolated causal roles and don't enter into our everyday mental lives: they are not available for explicit decision-making and memory-formation processes, for example, and they can't control a wide range of behavior or become the objects of our introspection. The functional properties of conscious states, however, do not exhaust our everyday concept of consciousness. The causal interactions outlined above are normally accompanied by certain *experiential* qualities: conscious states,

in addition to their functional roles, 'feel' like something when we undergo them. Non-conscious states, on the other hand, aren't accompanied by any sort of feeling, sensation, or experiential quality.

The scientific work on consciousness outlined above focuses on the functional aspects of conscious states: their "cognitive accessibility" (Block, 2007) or their "availability for global control" (Chalmers, 1997). This leads to greater understanding of what conscious states do and how they do it, but doesn't seem to add to our understanding of the phenomenal properties of consciousness. (See Block, 1995, for further discussion of the distinction between these two aspects of consciousness.) What should we conclude from this? Perhaps these particular scientific experiments weren't trying to explain phenomenal consciousness, and other experiments could do a better job. Perhaps there aren't any good explanations of phenomenal consciousness now, but they will follow advances in science and technology. Or perhaps, as some philosophers think, there's something about the very *nature* of phenomenal consciousness that evades scientific explanation. In order to understand such a claim, it is necessary to first say more about the phenomenal properties of conscious experience.

3. Phenomenal consciousness

When one hears the piercing shriek of the alarm clock, smells freshly-brewed coffee, endures the pain of a piece of grit in one's eye, or enjoys the sense of relaxation at the end of a busy day, there is something it *feels* like to have each experience. Furthermore, the 'something' that it feels like is peculiar to the kind of experience one is having – the visual experiencing of seeing a rainbow, for example, has a different kind of feeling from the auditory experience of hearing fingernails scrape down a chalkboard. It is this aspect of consciousness that has proved particularly fascinating to philosophers: the phenomenal character of conscious experience. (Philosophers also often use the term 'qualia' to refer to the qualities that make up the phenomenal character of the experience.) Phenomenal consciousness has a number of interesting features: it is *subjective*, it is seemingly *private*, and we have some form of *special access* to our phenomenally conscious states. This section covers each of these features in turn.

One way to understand what we mean by the subjectivity of conscious experience is through terminology introduced by Nagel (1974), who suggests that when you are undergoing a conscious experience, there is *something it is like* to be you:

> [F]undamentally an organism has conscious mental states if and only if there is something that it is like to *be* that organism – something it is like *for* the organism. We may call this the subjective character of experience. (p. 436, italics in original)

There is something it is like for me to see the color red, and there is something it is like for me to hear the sound of a bell (and *what* it is like differs across the two experiences.) There is presumably nothing it is like to be a rock or a piece of furniture. What about non-human animals? We tend to assume that other mammals, at least,

experience a world of sights and sounds. But not all mammals have visual and auditory systems like our own: consider the bat, which uses echolocation instead of vision to build up information about the world:

> [B]at sonar, though clearly a form of perception, is not similar in its operation to any sense that we possess, and there is no reason to suppose that it is subjectively like anything we can experience or imagine. (Nagel, 1974, p. 438)

Nagel asks us to think about *what it is like* to be a bat: what sort of feeling or sensory quality accompanies the echolocation process? Nagel's point is that no matter how much we study the bat's physiology, we don't get close to understanding what the bat's subjective experiences are like. This poses a problem for the attempt to provide a science of consciousness:

> If we acknowledge that a physical theory of mind must account for the subjective character of experience, we must admit that no presently available conception gives us a clue how this could be done. (Nagel, 1974, p. 445)

Part of the problem doubtless lies in the limits of our current technology, and the restrictions governing ethical experimentation. Functional neuroimaging technology, for example, does not allow us direct access to people's thoughts. Instead, it measures the amount of oxygen in the blood across different regions of the brain, which can be used as evidence of cognitive activity only after interpretation. There are more direct ways of measuring the brain's relation to conscious experience, but these involve invasive experiments on living brains, such as single-cell recordings, which are usually only carried out on non-human animals. The neural data from such experiments are more precise than in neuroimaging, but how can we establish with which conscious experiences they are correlated? These experiments are restricted to non-human animals who cannot verbally describe their experiences as humans can.

Technological advances however, are being made. For example, human and animal behavioral, electrophysiological and brain-imaging studies of visual consciousness during binocular rivalry are summarized by Miller (this volume), and described in detail in the companion volume (Miller, 2013). But even if still further technological and experimental advances occur, will we be able to identify and measure conscious experiences? There remains the problem that phenomenal consciousness seems markedly different from the other sorts of data that scientists study: not only is it subjective, it is also seemingly *private*. You might easily be able to imagine the experiences of other human beings (in the way you can't imagine what it's like to be a bat) but verification is a different matter. Were you and I to experience a sunset together, you would have no way of knowing whether my color experiences – what it was like for me – were the same as your own. We can use the same words to describe our experiences, but this doesn't guarantee that the experiences themselves have the same sort of subjective quality for each of us. Science deals in data that are public in the sense of being shareable and measurable by others; the privacy of conscious experiences seems to prevent them fulfilling these criteria.

Of course, there is a wealth of public data *associated* with our conscious states: psychological data such as our verbal reports of our conscious states, and our behavioral discriminations between different kinds of stimulus; and neurological data such as those from electroencephalography and neuroimaging. Our ability to discriminate between different colors can be measured, for example, as in color-blindness tests. And when someone is undergoing a migraine headache, the chemical composition of their neurotransmitters can be measured. But measuring a person's ability to discriminate colors doesn't establish *what it is like* for them to experience those colors, and measuring the bodily activity that correlates with the pain is not the same as measuring the feeling of the pain itself for the person undergoing it. The privacy of conscious experience makes it markedly different from standard scientific data.

Given that psychology and neuroscience *can* measure the neural and behavioral markers associated with phenomenal consciousness, it might seem tempting to simply treat the conscious experience as nothing more than its measurable markers. For example, if a person's experience of pain is reliably correlated with a particular pattern of blood-oxygenation in the brain, then we might feel tempted to say that to record that particular pattern of neural activity just *is* to record the experience of pain. This would be to 'operationalize' the concept of phenomenal consciousness, and identify the experience with the markers of that experience. While this would make phenomenal consciousness a less mysterious subject of scientific study, there are good reasons to resist this move. To see why it's important to retain the distinction between the experience itself and the experience-related behavior, consider the asymmetry between our knowledge of our own conscious states and those of others. When I'm in pain, a doctor can know I'm in pain on the evidence of my pain-related behavior: she infers that I'm in pain from my verbal reports, my bodily injuries, my neural activity. All of these count as evidence for the doctor that I am in pain. But notice that I don't need any of these pieces of evidence to know that *I* am in pain. The existence of my pain isn't something I infer from anything else: when I know that I'm in pain, I know it directly, in a way that doesn't seem accessible to anyone else. It doesn't even seem to make sense to ask me what my evidence is. I am simply directly aware of my pain. There is an asymmetry, then, between the (non-inferential, direct) way I know my own conscious states and the (inferential, evidence-based) way that others know my conscious states.

This asymmetry only applies to our conscious states, and not our other internal states: when it comes to the results of a blood test, for example, the doctor and I both base our knowledge on the same evidential data. If I disagree with the doctor about the results, our disagreement will ultimately be settled by further public data; I don't have any special access to the properties of my blood that the doctor lacks. But such an appeal to the evidence wouldn't settle the issue if I was disagreeing with a doctor about whether I'm in pain, because my first-person knowledge of my conscious experience is a different sort of knowledge from that provided by third-person scientific evidence. Furthermore, it's hard to see how the third-person scientific evidence could 'trump' my first person knowledge of my experience: what sort of evidence could a

doctor use to persuade you that you're in pain, when you are awake and conscious but not consciously experiencing pain?

The 'special access' that we have to our conscious states goes along with their subjectivity and privacy to create a problem for our scientific understanding of consciousness. Science deals with data that is objective, measurable, and knowable to different people via the same evidence; phenomenal consciousness seems not to meet these requirements. If we want to have a scientific understanding of consciousness, it looks like we need an explanation of the first-person nature of conscious experience in third-person terms. But to expect a scientific explanation of consciousness is already to assume that consciousness is part of the physical world that scientists study – and this assumption is a matter of philosophical debate, as the following section will discuss.

4. Is consciousness physical?

Scientific research helps us achieve an understanding of the physical world of particles, charges, enzymes, cells, organisms, and so on. We can only have a scientific understanding of consciousness, therefore, if conscious experiences are part of the physical world. And there is good reason to think that consciousness is closely connected to the physical brain: we know that changes to the brain, such as those caused by head injury or by toxins in the bloodstream, can result in changes to conscious experiences; and brain scans performed on meditating subjects show their altered conscious states correlating with differences in their neural activity. Such systematic and reproducible effects persuade many of the truth of *physicalism*, the view that consciousness is part of the physical world. But there are reasons to doubt physicalism. If conscious experiences are just brain states, then why have they proved so elusive to scientific explanation? And how could a physical thing like a brain state have a 'phenomenal feel'?

If we had a complete description of our world in physical terms, physicalism and non-physicalism disagree about what would follow from this description. According to physicalism, once the physical facts are fixed then the facts about conscious experience are also fixed: from a complete physical description of the world, facts about consciousness would follow. Non-physicalism claims that even once the full physical description of the world is fixed, facts about conscious experience are further facts which are not directly entailed by the physical facts. This section will examine both views.

4.1 Physicalism

The claim that consciousness is part of the physical world can be understood in different ways. One obvious way is to claim that each type of conscious experience is *identical* with a certain type of brain event: to experience the smell of fresh coffee, for example, is just to have a particular pattern of neural firing in a certain brain area. Proponents of this so-called 'type-physicalism' include Smart (1959) and more

recently Polger (2004). Notice that *identifying* a type of conscious experience with a type of neural event entails that anyone lacking that type of neural event also lacks the relevant type of conscious experience. Imagine we come to identify the sensation of pain with a certain kind of neural firing, and then discover a creature – human or otherwise – that lacks the appropriate kind of neural firing. According to type-physicalism, that creature cannot be in pain, no matter what behavioral evidence we have to the contrary. And consider the case of creatures, such as octopuses, that have very different nervous systems from humans; if there aren't any kinds of neural event that humans and octopuses share, we'd have to deny that octopuses could ever be in pain. Notice that the opposite also holds: if we find a creature that has the appropriate pattern of neural activity, we'd have to conclude it was in pain, even if it demonstrated none of the behavioral signs.

A common way to avoid these problems is to adopt 'token-physicalism' about consciousness. Token-physicalism is a weaker view than type-physicalism, because it makes claims only about particular instances ('tokens') of conscious experiences. While type-physicalism claims that each *type* of conscious experience is identical with a *type* of neural event, token-physicalism claims merely that each instance of a particular conscious experience is identical with some neural event or another. Token-physicalism allows that my pain experience is identical with a neural firing (and therefore physical), and that another creature's pain experience is identical with a neural firing (and therefore physical), without those neural firings being of the same kind. This picture allows for the 'multiple realizability' of conscious mental states: the idea that one kind of conscious experience can be physically realized in multiple ways (Putnam, 1967).

While token-physicalism allows us to retain the claim that conscious experiences are physical things, it lacks the scientific usefulness of type-physicalism. Type-physicalism, if true, would allow us to make predictions and generalizations: if a type of experience is identical with a type of neural event, we know that other creatures with that kind of neural event have the same kind of experience, and vice versa. Token-physicalism, however, leaves us seeking explanations. In virtue of what, for example, do two distinct kinds of neural event realize the same kind of experience? Philosophers have tried to remedy this situation by finding generalizations at a more abstract level, for instance by claiming that two types of neural event realize the same type of conscious experience in virtue of playing the same *functional role*.

According to this 'functionalist' approach, the neural event identical with my pain and the neural event identical with another creature's pain are playing the same functional role: each is caused by physical injury, and leads one to believe that one is in pain, and to the desire to be pain-free, for example. Although the neural events themselves are of different neural types, they are both realizers of the same functional type: pain. Functionalism works reasonably well as a theory of mental states like beliefs and desires, but is less convincing when it comes to capturing the qualitative character of conscious experiences. To see why, notice that the functionalist claims that any system with the same functional organization as you will have all the same mental states as

you. And then consider what would happen if we took the functional organization of your one-billion neurons and implemented the functional roles in something other than your brain:

> Suppose we convert the government of China to functionalism, and we convince its officials to realize a human mind for an hour. We provide each of the billion people in China (I chose China because it has a billion inhabitants) with a specially designed two-way radio that connects them in the appropriate way to other persons and to [an] artificial body […] It is not at all obvious that the China-body system is physically impossible. It could be functionally equivalent to you for a short time, say an hour. (Block, 1978, p. 279)

Our intuitive response to Block's thought-experiment is to deny that the nation of China has any mental states at all, suggesting that functional role is an insufficient characterization of mentality. But in particular, it is hard to see how the China-body system could have the phenomenal properties of conscious experience:

> there is prima facie doubt whether it has any mental states at all – especially whether it has what philosophers have variously called "qualitative states," "raw feels," or "immediate phenomenological qualities." […] In Nagel's terms (1974), there is a prima facie doubt whether there is anything which it is like to be the [China-body] system. (Block, 1978, p. 281)

These problems for both type- and token-physicalism need not deter the physicalist, as the most minimal form of physicalism requires neither. Minimal physicalism about consciousness requires only that conscious experience *supervenes* on the physical world. To understand the concept of supervenience, consider David Lewis's (1986) example of a dot-matrix image:

> A dot-matrix picture has global properties – it is symmetrical, it is cluttered, and whatnot – and yet all there is to the picture is dots and non-dots at each point of the matrix. The global properties are nothing but patterns in the dots. They supervene: no two pictures could differ in their global properties without differing, somewhere, in whether there is or there isn't a dot. (p. 14)

The point Lewis is making is that there is nothing more to the picture than the dots: any identical arrangement of dots will yield the same patterns, because the patterns supervene on the dots. By analogy, if consciousness supervenes on the physical, then there is nothing more to it than the underlying physical arrangement of the world: no two identical physical worlds could differ with regard to consciousness. (For a more detailed discussion of the varieties of supervenience, see Kozuch & Kriegel, this volume.)

Approaching consciousness in terms of supervenience gives us a minimal form of physicalism, but one which is without explanatory power. To say the dot-matrix pattern supervenes on the dots does not tell us why those particular patterns exist, or how we should get clearer on the relationship between the dots and the patterns.

Similarly, supervenience physicalism doesn't tell us *how* consciousness arises from the physical world: it merely states that there is a co-variation between the world's physical properties and its conscious properties. Despite its minimalist nature, however, supervenience physicalism is still open to objections from the non-physicalist.

4.2 Non-physicalism

The non-physicalist does not have to deny that the physical world plays an important role in generating our mental states. They merely have to deny the supervenience physicalist's claim that physical facts about neural activity, for example, fix the existence and nature of conscious experience. Non-physicalists have developed several thought experiments to persuade us of the falsity of physicalism, the most famous of which are Jackson's (1982) 'knowledge argument' and Chalmers' (1996) 'zombie argument'.

Frank Jackson's knowledge argument introduces the character of Mary. Jackson asks us to imagine that Mary has been brought up from birth in an entirely black-and-white room. Mary has grown up reading black-and-white books and watching black-and-white television, and has developed a vast scientific knowledge as a result. In particular, she has learned how color vision works in humans: she knows everything that science can tell her about light wavelengths, for example, and visual circuitry in the brain. Suppose that after many years of study in her black-and-white room, Mary has come to know all the physical facts about human color vision. What will happen when Mary leaves the black-and-white room and enters the world of color for the first time? When Mary sees something colored red for the first time, for example, will she learn something new?

> It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had *all* the physical information. *Ergo* there is more to have than that, and Physicalism is false. (Jackson, 1982, p. 130, italics in original)

Most people share the intuition that Mary will learn something new: she will learn what it's like to experience the color red. But because the thought experiment stipulates that Mary had already learned all the *physical* facts about color in her black-and-white room, Mary must be learning something *non-physical* when she learns what it is like to see red. The strength of the knowledge argument, as Jackson realizes, is that cases like Mary's compel us to think that "one can have all the physical information without having all the information there is to have" (Jackson, 1982, p. 130).

The knowledge argument has created a great deal of literature, mostly constituted by attempts to defend physicalism by challenging the thought experiment. Some philosophers claim that Mary could not have all the physical facts in the first place (Alter, 1998), while others argue that Mary's new knowledge is made true by physical facts she already knew (Lycan, 1996). Another physicalist tactic involves denying that Mary acquires new factual knowledge at all: instead she has perhaps gained a set of abilities

(Lewis, 1983) or 'become acquainted' with color experience (Conee, 1994). (For a more detailed discussion of these options, see Brogaard, this volume.)

David Chalmers' zombie argument asks you to imagine an organism that is physically identical to you, but which lacks the phenomenal experiences that you have: call this a 'zombie'. Zombies behave just like human beings. When a zombie cuts itself, for example, it bleeds, it emits sounds of pain, and certain areas of its brain are activated (and these are the same areas that are activated in your own brain when you cut yourself). The only difference is that there is not "something it is like" to be a zombie: the zombie does not experience any painful sensations. Chalmers argues that the very fact we can imagine zombies fitting this description entails that physicalism is false. To see why, remember that physicalism takes the facts about conscious experience to be determined by the physical facts. If physicalism is true, therefore, it would be *impossible* for a world physically identical to ours to differ in terms of phenomenal consciousness. This means that in order to prove physicalism false, it is enough to show the *possibility* of a world physically identical to ours that differs in its phenomenally conscious properties. A world containing zombies would be such a world: if zombies are possible, physicalism is false. But does *imagining* zombies show that they are *possible*? Chalmers thinks that zombies are conceivable, where conceivability amounts to "ideal conceivability, or conceivability on ideal rational reflection" (Chalmers, 1999, p. 47): there is no contradiction, he claims, in the idea of zombies. While some physicalists respond by denying that zombies are actually conceivable (e.g., Dennett, 1995), the more common physicalist response to the zombie argument is to deny that the conceivability of zombies entails their possibility (see e.g., Hill & McLaughlin, 1999; Loar, 1990; Yablo, 1993). Chalmers (1996, 1999) responds to the physicalist by characterizing the notions of conceivability and possibility in the formal framework of two-dimensional semantics, and arguing that claims about conceivability entail claims about possibility when we isolate the relevant meanings of 'conceivable' and 'possible'. (The 'zombie' argument is discussed at greater length by Brogaard, this volume.)

One of the most famous historical arguments against physicalism by Descartes also makes claims about what is possible on the basis of claims about what we can conceive. In his Sixth Meditation, Descartes suggests that he can conceive of his mind existing without his body, therefore it is not possible for his mind to be dependent upon his body:

> although I certainly do possess a body with which I am very closely conjoined; nevertheless, because, on the one hand, I have a clear and distinct idea of myself, in as far as I am only a thinking and unextended thing, and as, on the other hand, I possess a distinct idea of body, in as far as it is only an extended and unthinking thing, it is certain that I, [that is, my mind, by which I am what I am], is entirely and truly distinct from my body, and may exist without it.
>
> (Descartes, 1641/1996, p. 107)

This leads Descartes to conclude that the mind is an entirely distinct substance from the body. His view, that the physical world of measurable and locatable objects is

supplemented by a distinct realm of mental substance, has become known as 'substance dualism'. It is important to notice that not all arguments against physicalism lead to substance dualism. While substance dualism is committed to the idea that the mind could exist without the physical body, the knowledge argument and the zombie argument, for example, are intended only to establish that the physical facts don't fix the facts about phenomenal consciousness. As a result, the knowledge argument and the zombie argument can only be used to establish that there are non-physical *properties*, not (without amendment) that there are non-physical *substances*. The resulting 'property dualism', unlike substance dualism, is committed to the claim that all substance is physical. However, property dualists (e.g., Kim, 2005) suggest that at least some things have non-physical properties in addition to their physical properties. In the case of human brains, for example, the property dualist might think that neural matter can have non-physical properties (e.g., subjective feel) in addition to physical properties (e.g., firing rate). Property dualism is consistent with some forms of physicalism. If, for example, we interpret 'token-physicalism' as the claim that every conscious state has physical properties, then it looks like one can be a token-physicalist *and* a property dualist: having physical properties is consistent with also having non-physical properties. However, the relationship between property dualism and physicalism is a complex one, and Schneider (2012) has argued that property dualists must reject physicalism altogether. Property dualism, she argues, "does not lend itself to any physicalism worth having" (Schneider, 2012, p. 62).

Both substance and property dualism face problems when it comes to accounting for causal interactions between the mind and the rest of the world. Our conscious states seem to be causally integrated with aspects of the physical world: my feeling of embarrassment (conscious state) causes me to blush (physical state); my low levels of blood-sugar (physical state) cause me to feel dizzy (conscious state). If substance dualism is correct, then these events require physical and non-physical substances to interact with each other. Our knowledge of causes and effects, however, comes from observing interactions between wholly physical objects, and it's difficult to understand how a non-physical substance could participate in causal transactions. Property dualism doesn't face the same problem, as it denies the existence of non-physical substances. The property dualist can claim that conscious states are physical states, and thus there is no mystery about how they can cause physical events. But there is still a residual problem here for the property dualist: in virtue of which of its properties does the conscious state play that causal role? If it is in virtue of its non-physical properties, we are back to a version of the problem for substance dualism: it seems utterly mysterious how non-physical properties can interact with physical properties. But if the conscious state plays its causal role in virtue of its physical properties, then it looks like the (non-physical) conscious feeling of embarrassment is not what is causing the blushing. Without the physical properties, any causal interactions seem mysterious; but as soon as we have physical properties doing the causal work, the non-physical properties seem to be irrelevant. This is problem of 'causal exclusion' (Kim, 1989).

Options for solving the problem include allowing events to have both a physical and non-physical cause ('overdetermination') or denying that conscious properties are causal at all ('epiphenomenalism'). (See Bennett, 2007, for discussion of the various problems surrounding mental causation.)

5. Theories of consciousness

To take a stance in the debate over physicalism versus non-physicalism is to take a stance on the metaphysical nature of consciousness. It is not yet to provide a *theory* of consciousness: to say that consciousness is physical is not to commit to any particular view of what consciousness is, how it arises, or why it exists.

The main strategy of philosophers attempting to explain the phenomenal properties of conscious states has been to analyse consciousness in terms of *representation*. The notion of representation already features in the literature on mental states more generally, as philosophers use representation to make sense of how beliefs, desires, and other thoughts can be *about* certain things or states of affairs. To believe that Paris is in France or to desire that the rain stays off is to be in a representational state with a certain content: the content *that Paris is in France* or *that the rain stays off*. Philosophers have used this concept of representation in two different ways to account for phenomenal consciousness. The first way involves understanding phenomenal properties as straightforward 'first-order' representational properties, while the second way involves understanding phenomenal properties as 'higher-order' representational properties: representations of representations.

The first way to understand phenomenal properties in terms of representations is simply to think that the phenomenal properties of a conscious state are just its representational properties. Experiences, like beliefs and desires, can have representational contents. When I see a red balloon or hear a trumpet, I am having a perceptual experience with a representational content. The 'representationalist' about consciousness (e.g., Dretske, 1995; Tye, 1994) wants to say that the qualitative properties of the experience (what it is like to see red, to hear the trumpet) are just properties of the representation. It is true that I can have a belief about a red balloon or a desire to hear the trumpet without any accompanying qualitative features, but when this same representational content appears in a perceptual state, we *experience* some of those representational properties.

One problem faced by the representationalist approach to consciousness is that, while it works well for visual experiences, it is more difficult to see how it applies to other perceptual modalities. How should we characterize the representational contents of smells and tastes, for example? The problem becomes more obvious when we consider non-perceptual conscious states, like pains and moods: it seems to make sense to think of a perceptual experience representing the world in a certain way, but it is less clear how pains and moods can be understood as representing anything.

The second approach to understanding phenomenal consciousness in terms of representations focuses on 'higher-order' representations. Higher-order theories of consciousness start from the assumption that we have lots of representational mental states, not all of them conscious. They claim that phenomenally conscious mental states are those representational mental states that are (or are disposed to be) the object of a further representational state: a higher-order representation. An unconscious mental state becomes a conscious mental state, according to higher-order theories, when we reflect or focus on it. The exact nature of this 'reflecting' or 'focusing' depends on whether we think of the higher-order mental state as a kind of perception, or as a kind of belief. On the former view (e.g., Lycan, 1996), mental states are consciously experienced when they are perceived or sensed in the right way; while on the latter view (e.g., Rosenthal, 1986), mental states are consciously experienced when they become the object of a belief-like thought.

One problem for higher-order theories of consciousness is that they posit fairly extensive cognitive mechanisms and capacities. This is a problem for all versions of the view, but particularly for those who focus on higher-order thought rather than perception. To represent one's own mental representations, and particularly to think about one's own thoughts, puts high demands on the cognitive architecture possessed by a creature that has conscious states. One might think that many non-human animals possess basic phenomenally conscious states, without possessing the capacity for thought at all, let alone higher-level thought. Higher-order views of consciousness seem to link the ability for basic conscious experience too tightly to complicated cognitive capacities.

Some theories of consciousness also attempt to explain *why* consciousness exists: what is it *for*? Given that our brains can perform all sorts of complex processes unconsciously, why do we consciously experience some of it? Going back to the functional properties of consciousness discussed in section two, recall that our conscious states interact in a certain way with other mental states: our conscious states can be verbally reported, stored in memory, introspected, used to guide action, and so on. One might think that the function of consciousness is to make certain information 'cognitively accessible' or 'globally available' to other parts of the cognitive system. The interactions between conscious states seem to enable the integration of lots of information from different areas of the cognitive system: sensory information, linguistic data, motivational concerns, and so on. This leads some theorists (e.g., Baars, 1988) to propose that the function of consciousness is to 'broadcast' information, by providing a 'global workspace' for sharing information.

All of these approaches to consciousness are 'top-down', in the sense that they start from our concept of consciousness and attempt to analyze it in terms of other notions such as representation, information, and function. They are all compatible with physicalism, but they are framed at a higher level of abstraction than the neural level: these theories show how a physicalist theory of consciousness might work without mentioning biochemical features of the brain. But once the theoretical details are in place, we

can look for the neural events that implement these abstract roles. Representationalist theories of consciousness, for example, can look to work in computational psychology and neuroscience that invoke semantically-interpreted and physically-implemented functional states. (For a neuroanatomical interpretation of representationalism, see Mehta & Mashour, 2013.) Higher-order theories of consciousness seem to require a 'metacognitive' monitoring process, and such processes are associated with the prefrontal cortex of the brain, so evidence for the neural implementation of higher-order theories might best be sought in prefrontal neural processing (Lau & Rosenthal, 2011). And the global workspace theory of consciousness has also been developed as a global *neuronal* workspace theory, using computer modeling of neural networks to simulate brain activity (Dehaene, 2001).

The motivation behind all of these top-down approaches to consciousness is that consciousness arises in the human brain in virtue of its organization and structure: while we may seek the implementation of consciousness in neurons, the molecular properties of the brain are less important than its higher-level functioning. The alternative to top-down approaches involves starting from basic brain activity, and using neural features such as firing activity, chemical composition, or location to shed light on consciousness. These 'bottom-up' approaches to consciousness variously suggest that conscious experience might be the result of neural activity in a particular area of the cortex, for example, or the speed at which neurons fire, or the complex interaction of such features. Bottom-up approaches can differ with regard to the level of neural activity they focus on: some propose that consciousness might be the product of individual neurons, or of populations of neurons, or of whole brains. In the case of individual neurons, this might mean that a conscious experience of a certain feature is the product of a receptive-field neuron that responds to a particular sensory property, such as a particular frequency in the auditory field, or a particular direction of motion in the visual field. Alternatively, conscious experience might be the product of neural networks: conscious experience of a certain sort might depend on the outcome of a competition between neural coalitions, or on the synchrony of neural populations oscillating at a particular frequency. A further suggestion is that the source of consciousness is not single neurons or groups of neurons, but the brain as a whole: according to this view, we should be looking at global patterns of brain activity such as the resonance of an electromagnetic field, or quantum effects occurring in subcellular structures. (For examples of each of these approaches, see Kouider, 2009.)

Notice that one doesn't need to commit to physicalism about consciousness in order to be interested in the neurobiological research associated with consciousness. A non-physicalist can accept that there are correlations between neural features and conscious experiences, as long as they deny that these neural correlates of consciousness are sufficient to fix the facts about conscious experience.

6. Explaining consciousness

It's not always clear whether a theory of consciousness is attempting to explain everything about consciousness, or merely some aspect of conscious states. Some theories of consciousness focus solely on visual consciousness and say nothing about other sensory modalities, while others concentrate on the varieties of conscious experience associated with imagination or proprioception. Even those theories attempting to account for consciousness more generally tend to focus on explaining certain properties of conscious experience: functional properties like its integrational or introspectable properties, or structural properties like its unity or its temporal dynamics. What's less clear, however, is whether any theory of consciousness explains the *phenomenal* properties of conscious experience: *what it is like* to be having a conscious experience. Whenever a theory presents us with an analysis of consciousness, we always seem to be able to ask "But why does it *feel* like this?" There seems to be what Levine (1983) calls an "explanatory gap" between what explanations of consciousness actually tell us, and what we want them to tell us.

To understand this explanatory gap, consider a standard scientific explanation of heat as the motion of molecules. We understand why heat has the properties it does by identifying it as a certain kind of molecule, and we don't feel that something crucial has been left unexplained:

> It ["Heat is the motion of molecules"] is explanatory in the sense that our knowledge of chemistry and physics makes intelligible how it is that something like the motion of molecules could play the causal role we associate with heat. […] Once we understand how this causal role is carried out, there is nothing more we need to understand. (Levine, 1983, p. 357)

By contrast, notice what happens if we were to identify a certain kind of conscious experience with a certain physical state – seeing red as a pattern of activity in area V4 of the visual cortex, for example, or Levine's own example of pain as C-fibers firing. As in the heat case, the claim that "pain is the firing of C-fibers" is explanatory insofar as it explains why the firing of C-fibers results in avoidance effects or verbal responses: it tells us that the causal role of pain is being played by C-fibers. But explaining the causal role seems insufficient, as Levine (1983) points out:

> However, there is more to our concept of pain than its causal role, there is its qualitative character, how it feels; and what is left unexplained by the discovery of C-fiber firing is *why pain should feel the way it does!* […] the identification of the qualitative side of pain with C-fiber firing (or some property of C-fiber firing) leaves the connection between it and what we identify with it completely mysterious. (p. 357, italics in original)

The problem is that by identifying consciousness with a physical state (or a functional state), we're left with an incomplete explanation. The explanation fails to make it intelligible to us why conscious states feel the way they do, or indeed why they feel like anything at all.

The claim that there is an explanatory gap in our explanations of phenomenal consciousness is an epistemological claim: a claim about how we understand (or fail to understand) the world, rather than a metaphysical claim about how the world is independently of our understanding. But for a physicalist, the explanatory gap poses a particular challenge. If phenomenal consciousness is part of the physical world, as the physicalist claims, why do we fail to capture it with our explanations? One physicalist response is to simply deny the existence of an explanatory gap, or claim that it is only an *apparent* explanatory gap. The fact that we can imagine a sensation other than pain, or no sensation at all, accompanying the firing of C-fibers doesn't mean that C-fibers are not an adequate explanation of pain. To think that there is something missing from the explanation is to be misguided, according to deniers of the explanatory gap. A second physicalist response is to accept that there is currently an explanatory gap between what we know about the brain and the phenomenal properties of consciousness we want to explain, but claim that this is due to the limits of current scientific knowledge. On this view, the explanatory gap will be closed when we make further progress on brain science. A third way for the physicalist to respond is to suggest that the explanatory gap is due not to our limited current science, but to the way we understand consciousness. Perhaps our concept of consciousness leads to the explanatory gap: consciousness itself is a physical phenomenon, but the way we think about consciousness leads us to find our physical explanations unsatisfactory. These three responses to the explanatory gap all constitute attempts to save physicalism: they claim that if there is a gap, it's a gap in our knowledge or understanding. The alternative view, proposed by the non-physicalist, is that the explanatory gap exists because there really is a 'gap' in the world. In other words, there is a gap between physical descriptions of the world and phenomenal consciousness because phenomenal consciousness cannot be captured by any physical description of the world. We cannot give an adequate physical explanation of consciousness, according to this position, precisely because consciousness is not part of the physical world. Sometimes this argument is put forward as an inference to the best explanation for the explanatory gap, but there is a stronger argument available. Chalmers and Jackson (2001) argue that the existence of a genuine explanatory gap would be proof that consciousness is not physical: if consciousness were physical, then we would be able to explain it.

Sometimes when we have trouble explaining a certain phenomenon, it's because we are mistaken about the nature of the phenomenon we are trying to explain. In pre-Copernican times, for example, people tried to explain the movement of the sun around the earth and ended up with inadequate explanations. It turned out that what they actually wanted to explain was the movement of the earth around the sun, which was responsible for the apparent movement of the sun around the earth. Examples like these suggest that perhaps we find an explanatory gap between our physical explanations and phenomenal consciousness because we're wrong about the nature of phenomenal consciousness. But it's not clear that we can be wrong about phenomenal consciousness in the way people were wrong about the sun's movement around the earth. In the latter case, aspects of the world turned out to be different from how

people experienced them. In the case of phenomenal consciousness, however, we'd be committed to saying that aspects of our experience are not how we experience them to be. Can we even make sense of this notion? It seems to require saying that we could be *wrong* about what it is like to see colors or hear sounds, and that we could be *wrong* about our own pain sensations. The very concept of phenomenal consciousness, as we saw earlier, seems to involve a kind of subjectivity and special access that make it difficult to understand how we could be wrong.

Many of the phenomena studied by scientists and philosophers are highly technical and seem far-removed from everyday life. Consider a physicist exploring the behavior of fundamental particles, for example, or a metaphysician pondering the relation between facts and propositions. But the topic of *consciousness* seems to be different, in the sense that we're already familiar with consciousness through our everyday experiences without requiring special training or expertise. Long before we're capable of learning from textbooks, we know what it is like to experience sights and sounds, and pains and pleasures. And the way we know about our conscious experiences is 'from the inside', in a way that is difficult to share or to measure. These peculiarities of phenomenal consciousness pose problems for both scientists and philosophers in their attempts to reconcile our everyday view of conscious experience with our scientific knowledge of the brain.

7. Conclusion

Scientific research on consciousness has made, and continues to make, significant breakthroughs. Some aspects of conscious experience, however, render it more difficult to approach scientifically: our experiences seem to have subjective and qualitative features that elude scientific techniques of experimentation and explanation. This prompts philosophical questions regarding the nature of conscious experience itself, and its place in the physical world. Such concerns, however, should not hinder progress in the science of consciousness: science can provide information about the neural activity that correlates with conscious experience, without making claims about the kind of dependency relations (identity, supervenience, etc.) that ground these correlations (see Hohwy & Bayne, this volume). But any theory of consciousness faces the question of how to account for the phenomenal properties of consciousness: how to bridge the 'explanatory gap' between the physical world and the features of experience. It is this task that Chalmers (1995) has labelled the 'hard problem' of consciousness:

> It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does. If any problem qualifies as *the* problem of consciousness, it is this one. (p. 201, italics in original)

References

Alter, T. (1998). A limited defense of the knowledge argument. *Philosophical Studies*, 90(1), 35–56. DOI: 10.1023/A:1004290020847

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge, MA: Cambridge University Press.

Bennett, K. (2007). Mental causation. *Philosophy Compass*, 2(2), 316–337. DOI: 10.1111/j.1747-9991.2007.00063.x

Block, N. (1978). Troubles with functionalism. In C. W. Savage (Ed.), *Minnesota studies in the philosophy of science*, Vol. 9, Perception and cognition: Issues in the foundations of psychology (pp. 261–325). Minneapolis, MN: University of Minnesota Press.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain and Sciences*, 18(2), 227–247. DOI: 10.1017/S0140525X00038188

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, 30(5–6), 481–499.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford, UK: Oxford University Press.

Chalmers, D. J. (1997). Availability: The cognitive basis of experience? In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates* (pp. 421–424). Cambridge, MA: MIT Press.

Chalmers, D. J. (1999). Materialism and the metaphysics of modality. *Philosophy and Phenomenological Research*, 59(2), 473–496. DOI: 10.2307/2653685

Chalmers, D. J., & Jackson, F. (2001). Conceptual analysis and reductive explanation. *Philosophical Review*, 110(3), 315–361. DOI: 10.1215/00318108-110-3-315

Conee, E. (1994). Phenomenal knowledge. *Australasian Journal of Philosophy*, 72(2), 136–150. DOI: 10.1080/00048409412345971

Dehaene, S. (2001). *The cognitive neuroscience of consciousness*. Cambridge, MA: MIT Press.

Dennett, D. C. (1995). The unimagined preposterousness of zombies. *Journal of Consciousness Studies*, 2(4), 322–326.

Descartes, R. (1641/1996). *Meditations on first philosophy. With selections from the objections and replies* (J. Cottingham, Trans., Rev. ed.). Cambridge, UK: Cambridge University Press.

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

Hill, C. S., & McLaughlin, B. P. (1999). There are fewer things in reality than are dreamt of in Chalmers's philosophy. *Philosophy and Phenomenological Research*, 59(2), 445–454. DOI: 10.2307/2653682

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32(127), 127–136. DOI: 10.2307/2960077

Kim, J. (1989). The myth of non-reductive materialism. *Proceedings and Addresses of the American Philosophical Association*, 63(3), 31–47. DOI: 10.2307/3130081

Kim, J. (2005). *Physicalism, or something near enough*. Princeton, NJ: Princeton University Press.

Kouider, S. (2009) Neurobiological theories of consciousness. In W. P. Banks (Ed.), *Encyclopedia of consciousness*, Vol. 2 (pp. 87–100). Oxford, UK: Elsevier. DOI: 10.1016/B978-012373873-8.00055-4

Lau, H. C., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15(8), 365–373. DOI: 10.1016/j.tics.2011.05.009

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64(4), 354–361.

Lewis, D. (Ed.). (1983). Postscript to "Mad pain and martian pain". In *Philosophical Papers*, Vol. 1 (pp. 130–132). Oxford, UK: Oxford University Press.

Lewis, D. K. (1986). *On the plurality of worlds*. Oxford, UK: Blackwell Publishers.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1), 6–14. DOI: 10.1016/S1364-6613(99)01418-7

Loar, B. (1990). Phenomenal states. *Philosophical Perspectives*, 4, 81–108. DOI: 10.2307/2214188

Lycan, W. G. (1996). *Consciousness and experience*. Cambridge, MA: MIT Press.

Mehta, N., & Mashour, G. A. (2013). General and specific consciousness: A first-order representationalist approach. *Frontiers in Psychology*, 4, 407. DOI: 10.3389/fpsyg.2013.00407

Miller, S. M. (Ed.). (2013). *The constitution of visual consciousness: Lessons from binocular rivalry*. Advances in Consciousness Research (Vol. 90). Amsterdam, The Netherlands: John Benjamins Publishing Company. DOI: 10.1075/aicr.90

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83(4), 435–450. DOI: 10.2307/2183914

Panconesi, A. (2008). Serotonin and migraine: A reconsideration of the central theory. *Journal of Headache and Pain*, 9(5), 267–276. DOI: 10.1007/s10194-008-0058-2

Polger, T. W. (2004). *Natural minds*. Cambridge, MA: MIT Press.

Price, N. S. C. (2013). Overview of visual system structure and function. In S. M. Miller (Ed.), *The constitution of visual consciousness: Lessons from binocular rivalry* (pp. 37–76). Advances in Consciousness Research (Vol. 90). Amsterdam, The Netherlands: John Benjamins Publishing Company. DOI: 10.1075/aicr.90.03pri

Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 37–48). Pittsburgh, PA: Pittsburgh University Press.

Rajala, A. Z., Reininger, K. R., Lancaster, K. M., & Populin, L. C. (2010). Rhesus monkeys (*Macaca mulatta*) do recognize themselves in the mirror: Implications for the evolution of self-recognition. *PLoS ONE*, 5(9), e12865. DOI: 10.1371/journal.pone.0012865

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, 49(3), 329–359. DOI: 10.1007/BF00355521

Schneider, S. (2012). Why property dualists must reject substance physicalism. *Philosophical Studies*, 157(1), 61–76. DOI: 10.1007/s11098-010-9618-9

Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review*, 68(2), 141–156. DOI: 10.2307/2182164

Tye, M. (1994). Qualia, content, and the inverted spectrum. *Noûs*, 28(2), 159–183. DOI: 10.2307/2216047

Yablo, S. (1993). Is conceivability a guide to possibility? *Philosophy and Phenomenological Research*, 53(1), 1–42. DOI: 10.2307/2108052