elizabethirv@gmail.com
mark.sprevak@ed.ac.uk

Eliminativism about consciousness

Liz Irvine
University of Cardiff

Mark Sprevak
University of Edinburgh

8 August 2018

1   Introduction

In this chapter, we examine a radical philosophical position about consciousness: eliminativism. Eliminativists claim that consciousness does not exist and/or that talk of consciousness should be eliminated from serious science. These are strong positions to take, and require serious defence. To evaluate these positions, the chapter is structured as follows. In Section 2 we introduce the difference between entity eliminativism and discourse eliminativism and outline the typical strategies used to support each. Section 3 provides a brief overview of the kinds of consciousness we refer to throughout the chapter. Section 4 focuses on entity eliminativist arguments about consciousness: Dennett's classic eliminativist argument (4.1); a rebooted version of Dennett's argument (4.2); and recent arguments for 'illusionism' (4.3). In Section 5, we examine discourse eliminativist arguments about consciousness: methodological arguments from scientific behaviourism (5.1); arguments based on the empirical accessibility of phenomenal consciousness (5.2); and a stronger version of discourse eliminativism aimed at both phenomenal and access consciousness (5.3). In Section 6, we offer a brief conclusion.

2   Eliminativism

If you meet an eliminativist, the first question to ask is, 'What do you want to eliminate: *entities* or *talk* about entities?' For any given *X*, an eliminativist might

assert either or both of:

1. *X*s do not exist
2. We should stop engaging in *X*-talk, using *X*-concepts, or other practices ostensibly associated with *X* in serious science

We will call (1) *entity eliminativism* and (2) *discourse eliminativism*. Entity eliminativists claim that we should expel a specific entity from the catalogue of entities assumed to exist. This may be a matter of removing a particular individual from our ontology (e.g. Zeus), but it could also be a matter of removing a property (e.g. being phlogisticated), an event (e.g. spontaneous generation), a kind (e.g. ghosts), or a process (e.g. extrasensory perception). In contrast, a discourse eliminativist seeks to eliminate certain ways of talking, thinking, and acting from science (e.g. talk about, and practices that attempt to investigate, gods, being phlogisticated, spontaneous generation, ghosts, or extrasensory perception).[1]

Entity and discourse eliminativism are distinct but obviously not unrelated positions. Opinions regarding an entity's existence should not float free from what we say, think, and do in serious science. However, the relationship between the two is not so tight that one form of eliminativism can be inferred from the other as a matter of course.[2] Someone might endorse one form of eliminativism but not the other. An entity eliminativist might wish to eliminate some entity (e.g. atoms) but to preserve talk, thought, and practices associated with that entity in science. For example, Mach claimed that atoms do not exist but he argued that physicists should continue to engage in atomic talk, thought, and action for its predictive and heuristic benefits: the atom 'exists only in our understanding, and has for us only the value of a *memoria technica* or formula' (Mach, 1911, p. 49). Conversely, a

---

[1]In focusing on 'serious' science, the discourse eliminativist makes no claim about whether this or similar talk, thought, and practice should be eliminated from *other* aspects of human life. What might be unacceptable to serious science may be tolerated, or welcomed, in popularisations of science, folk tales, religious practice, jokes, or science fiction. The boundary between 'serious' science and other aspects of human enquiry is not sharply defined. For the purposes of this chapter, we do not attempt to define it. We simply identify 'serious' science as work currently recognised as such by the scientific community, in contrast to, say, popular exposition of scientific research, adaptation of that scientific research for other ends, or training that is merely propaedeutic to conducting scientific research.

[2]Quine (1980) offered a bridge from discourse eliminativism to entity eliminativism with the quantificational criterion of ontological commitment. However, this bridge fails to link the two forms of eliminativism in a deductively certain way. It relies on numerous assumptions that are contentious in this context: assumptions about the aims of the scientific discourse, about the overriding importance of stating truth in science, and about the correct semantics for the discourse. Quine also only proposed his criterion for fundamental theories. Participants in this debate (realists and eliminativists about consciousness) are unlikely to agree about whether the theories in question are fundamental.

discourse eliminativist might wish to eliminate ways of talking, thinking, and acting in science but nonetheless say that the entity underlying this rejected discourse exists. In Section 5, we will see an example of this position with regard to scientific behaviourism's treatment of conscious experience.

Let us examine entity eliminativism and discourse eliminativism more closely.

Entity eliminativism is defined by its divergence from realism and agnosticism. A realist says, 'Xs exist', an agnostic says, 'We are not in a position to know whether Xs exist or not', and an eliminativist says, 'Xs do not exist'. In order for an entity eliminativist to have a genuine claim, she needs to have a genuine, not merely a verbal, disagreement with a realist and an agnostic. To this end, the eliminativist needs to make assumptions about what Xs are and those assumptions need to be shared with the realist and the agnostic. The realist, agnostic, and eliminativist should agree on what Xs *would be like if they were to exist*. What they disagree about is whether the entity *does* exist. Consequently, an argument for entity eliminativism generally involves two ingredients. The first is some way to identify the subject matter under dispute that is acceptable to all sides: realist, agnostic, and eliminativist. This is often done by providing a *description* of the essential properties of the entity, but, as we will see in Section 4.2, this is not the only way to do it. The second ingredient is an argument to show that no such entity exists. If the entity is identified by description, the second step may aim to show that no entity satisfies this description. Mallon et al. (2009) describe how this kind of argument has been used to support eliminativism about beliefs: first, claim that in order for something to be a belief, it must satisfy a certain description D (given in this case by folk psychology); second, argue that nothing satisfies description D (because folk psychology is false); third, since nothing satisfies D, conclude that beliefs do not exist.

In contrast to entity eliminativism, discourse eliminativism targets specific ways of talking, thinking, and acting in serious science. Let us say that the concept 'ding dong' refers to nanoscale, spherical, tentacled lifeforms. A discourse eliminativist about ding dongs says that scientists should stop talking, thinking, and pursuing research programmes about ding dongs. One motivation for doing this may be the conviction that there are no ding dongs (i.e. one is an entity eliminativist about them). But being an entity eliminativist about ding dongs is neither necessary nor sufficient for being a discourse eliminativist. One might think that ding dongs exist (or be an agnostic about them) but argue that scientists should avoid ding dong talk and thought because it is unproductive, misleading, or otherwise unhelpful. Conversely, one might think that ding dongs do not exist but argue that ding dong talk, thought, and practice is nonetheless useful to science and should be preserved: it may provide useful concepts (perhaps the ding dong concept is a useful way to group lifeforms) or encourage useful practices (looking for entities at certain spatial

scales). Even if ding dongs are not 'out there' in nature, they may be a useful fiction for science.

Arguments for discourse eliminativism typically consist of a negative part and a positive part. The negative part aims to establish that the talk, concepts, and practices targeted for elimination are somehow unhelpful, damaging, misleading, or otherwise problematic for science. In the case of conscious experience, a discourse eliminativist might argue that the relevant discourse is too subjective, hard to verify, does not generalise well, does not pick out a natural kind, produces intractable disagreements, does not cohere with other scientific talk, or otherwise leads to a degenerative scientific research programme. However, even if the discourse eliminativist's negative points land, they rarely suffice to achieve the discourse eliminativist's objective. Science seldom changes course unless a better alternative is already available. The positive part of a discourse eliminativist's argument aims to show that an alternative way of talking, thinking, and acting is available. The discourse eliminativist's says that this proposed alternative discourse is, on balance, better for achieving scientific goals than that targeted for elimination. Diverse virtues may weigh in this, including purely epistemic virtues (e.g. telling the truth, not positing things that do not exist) but also predictive, pragmatic, theoretical, and cognitive virtues.

3   Consciousness

In this chapter, we make use of Block's distinction between access consciousness and phenomenal consciousness (Block, 1990; Block, 2007). 'Access consciousness' refers to the aspects of consciousness associated with information processing: storage of information in working memory, planning, reporting, control of action, decision making, and so on. 'Phenomenal consciousness' refers to the subjective feelings and experiences that conscious agents enjoy: the feel of silk, the taste of raspberries, the sounds of birds singing, and so on. The latter is the 'feel-y', subjective, qualitative, what-it-is-like-ness, 'from the inside' aspect of consciousness. We use the term 'qualia' to refer to this subjective feeling.[3] Following Frankish (2016a), we define 'experience' in a purely functional way: mental states that are the direct output of the sensory system. This means that we do not assume that experience necessarily involves phenomenal consciousness.[4]

This chapter focuses on eliminativism about *phenomenal consciousness*. Access

---

[3]This is how Dennett uses 'qualia' but departs from the usage of some authors who take 'qualia' to refer to non-representational aspects of conscious experience.

[4]Other authors (including Block) equate phenomenal consciousness with experience. We adopt Frankish's usage here for (slightly better) ease of exposition.

consciousness will make an appearance in the final section (5.3). For the purposes of this chapter, we do not presuppose anything about the relationship between access and phenomenal consciousness, though some of the eliminativist arguments discussed below do take a stand on this.

One of the striking features of entity eliminativism about consciousness is that it is perceived as a philosophical position that is *self-evidently wrong*. Critics say that questioning the existence of phenomenal consciousness is impossible. Each of us knows, by introspection on our own case, that we have phenomenal consciousness – we know it for certain in a way that does not leave room for rational doubt. Not even Descartes doubted the existence of his subjective experience. Yet the eliminativist does this. Assessments of eliminativist claims have been correspondingly harsh. Frances writes, 'I assume that eliminativism about feelings really is crazy' (Frances, 2008, p. 241). Searle, 'Surely no sane person could deny the existence of feelings' (Searle, 1997). Strawson says that eliminativists 'seem to be out of their minds', their position is 'crazy, in a distinctively philosophical way' (Strawson, 1994, p. 101). Chalmers, 'This is the sort of thing that can only be done by a philosopher, or by someone else tying themselves in intellectual knots!' (Chalmers, 1996, p. 188). How can one deny this self-evident and foundational truth about our mental life? (As we will see in Section 4, entity eliminativism is typically combined with an attack against the reliability of our introspective access to phenomenal feelings.)

The discourse eliminativist about phenomenal consciousness faces a similar, though perhaps not quite so daunting, challenge. Discourse eliminativists identify problems with a scientific discourse and seek to offer a better alternative. The challenge faced by a discourse eliminativist about phenomenal consciousness is that phenomenal consciousness appears to be an incredibly important part of human mental life. Humans care about their phenomenal feelings: about the feelings that accompany eating their favourite dish, scoring a winning goal, being punched in the kidneys, or having their toes tickled. These feelings play a valuable, although hard to specify, role in our cognitive economy. For this reason, it seems that *some* reference to phenomenal consciousness should be made inside scientific psychology. A psychology that *never* talked about phenomenal consciousness would be incomplete in some way. Even if phenomenal feelings do not exist (as an entity eliminativist says), scientific psychology should still talk about phenomenal consciousness in order to explain why we (falsely, according to the entity eliminativist) take ourselves to be motivated by such feelings. Eliminating talk of phenomenal consciousness appears to ignore an important aspect of human mental life and would constitute a failure of ambition for scientific psychology.

4 Entity eliminativism about consciousness

In this section, we examine three entity eliminativist arguments about phenomenal consciousness. The first is Dennett's 'Quining qualia' argument (1988). The second is a rebooted version of Dennett's argument that aims to avoid the standard objection to that argument (namely, that Dennett mis-characterises phenomenal consciousness). The third is the recent research project of 'illusionism', which is related to Dennett's 'Quining qualia' argument but motivated on somewhat different grounds.

4.1 Dennett's eliminativism about qualia

Dennett's 'Quining qualia' paper appears to be a classic entity eliminativist argument: describe the essential properties of the alleged entity; show that nothing satisfies this description; conclude on this basis that no such entity exists. The identifying description that Dennett gives of phenomenal consciousness ('qualia') is that qualia are ineffable (not describable in words), intrinsic (non-relational), private (no inter-personal comparisons are possible), and directly accessible (via direct acquaintance). The final property is related to the idea that we have privileged, incorrigible, or infallible access to qualia. Dennett argues that no entity satisfies this description. As a result, 'Far better, tactically, to declare that there simply are no qualia at all' (Dennett, 1988, p. 44).

Dennett uses a number of 'intuition pumps' to get to this conclusion, which we summarise here.

First, it is plausible that how things subjectively feel to you is bound up with how you evaluate, or you are able to categorise or discriminate between, your experiences. Someone's first taste of a particular wine may be different to how it tastes to them after having become a wine aficionado. At first, my taste of the wine was bound up with judgements of yukkiness and an inability to easily tell one wine from another. My current taste of the wine is bound up with gustatory enjoyment and an ability to finely discriminate between different wines. This suggests that qualia are not intrinsic properties of experience: that the taste of this specific wine does not have a particular conscious, qualitative feeling (quale) for me independently of how I evaluate or categorise it. Instead, the way that wine (and other things) consciously tastes to me is at least partly determined by relational properties, such as whether I like it, and whether I can tell a Pinot Grigio from a Chardonnay.

This also puts pressure on qualia being directly accessible: it looks like I can't tell very much about my qualia from introspection. Say that as you get older, you start liking strong red wine more. One possibility is that your sensory organs have changed,

making strong reds taste different and more pleasurable compared to how they used to. On this scenario, you now have different qualitative feelings (more pleasant ones) on tasting strong reds to those you had before. Another possibility is that your sensory organs have stayed the same but your likings for specific experiences have changed. You have roughly the same conscious feelings as before but now you like those feelings more. On the first scenario, your qualia change; on the second scenario, your qualia stay the same. Dennett puts it to us that we would not be able to tell, merely from introspection, which scenario we are in. Yet this should be easy to do if we really had direct (or infallible or incorrigible) access to our qualia.

With respect to the putative ineffability and privacy of qualia, Dennett refers to Wittgensteinian arguments that render entirely private and incommunicable states senseless. Dennett goes on to argue that there is a way in which experiences are practically ineffable and private, just not in the 'special' way intended by qualiaphiles. Imagine two AI systems that learn about their environment in a fairly unsupervised manner and so develop different internal systems of categorising colour (adapted from Sloman and Chrisley, 2003). These two systems will end up with some states that are (at least to some degree) private and ineffable. One system's 'blue' states will be somewhat different to the other system's 'blue' states just in virtue of the internal differences in the systems (e.g. the 'blue' states of each system will be triggered by a slightly different range of hues). In the same way, humans can be in distinct (practically) ineffable and private states because of idiosyncrasies in their perceptual and cognitive processing. Some of these differences one may discover empirically (e.g. that you and I disagree about whether a particular paint chip is blue), and so we can make our experience more 'effable', and less private. Dennett's point is that ineffability and privacy of experience only amounts to this: practical and graded difficulties in assessing which internal state we are in, not essential properties of our experience.

In light of Dennett's considerations, it looks like supposedly phenomenal experience does not satisfy the description associated with qualia: there is nothing ineffable, intrinsic, private, and directly accessible to an experiencer that determines the way that things (phenomenally) seem to them. Whatever produces our judgements and reports about phenomenal experience, it is not an entity of the hypothesised kind. Qualia, as characterised by Dennett's description, do not exist.

A popular response to Dennett is to deny that qualia are accurately characterised by his description. This would effectively undermine his argument at the first step. Perhaps partly in response to Dennett, qualia realists nowadays tend to favour a more spare characterisation of qualia. Qualia need only have a *phenomenal character* (a 'what-is-it-like-ness' or subjective feeling) (Carruthers, 2000; Levine, 2001; Kind, 2001; Tye, 2002). They need not be intrinsic, private, ineffable, or directly accessible

properties of an experience. Frankish (2012) describes this as a strategic withdrawal by qualiaphiles from 'classic qualia' to 'diet qualia'.[5] Classic qualia are controversial posits (perhaps for the reasons Dennett gives); diet qualia are not:

> Philosophers often use the term 'qualia' to refer to the introspectively accessible properties of experiences that characterize what it is like to have them. In this standard, broad sense of the term, it is very difficult to deny that there are qualia. There is another, more restricted use of the term 'qualia', under which qualia are intrinsic, introspectively accessible, nonrepresentational qualities of experiences. In my view, there are no qualia, conceived of in this way. They are a philosophical myth. (Tye, 2002, p. 447)

Dennett does not wish to eliminate classic qualia only to retreat to realism about diet qualia. He wishes to eliminate *both* classic qualia and diet qualia.[6] However, we have seen that his argument is ineffective against diet qualia: the description does not pick out what qualiaphiles have in mind.[7] In the next section, we rework Dennett's eliminativist argument to target diet qualia.

4.2 'Quining qualia' rebooted

Dennett's 'Quining qualia' argument identified qualia by description. Once we switch to diet qualia, identification by description is ineffectual: there is no description that can identify diet qualia apart from their (contested) phenomenal feel. So, rather than identify the target for elimination (or realism) by description, we need to identify it some other way.

An alternative to identification by description is to identify qualia by a kind of ostension. The idea is to 'point' to alleged examples of qualia and then generalise to the kind they have in common. One asks one's interlocutor to consider those of her experiences that allegedly have qualia (consider the feel of silk, …), draw attention

---

[5]Frankish (2016a) describes a similar distinction between *weak* and *strong* illusionism, and Levine (2001) between *modest* and *bold* qualophilia.

[6]"Philosophers have adopted various names for the things in the beholder (or properties of the beholder) that have been supposed to provide a safe home for the colors and the rest of the properties that have been banished from the "external" world by the triumphs of physics: "raw feels," "sensa," "phenomenal qualities," "intrinsic properties of conscious experiences," "the qualitative content of mental states," and, of course, "qualia," the term I will use. There are subtle differences in how these terms have been defined, but I'm going to ride roughshod over them. In the previous chapter I seemed to be denying that there are any such properties, and for once what seems so is so. I am denying that there are any such properties' (Dennett, 1991, p. 372).

[7]Although we will not consider his reasoning here, Frankish (2012) argues that diet qualia and classic qualia are not distinct and so Dennett's 'Quining qualia' argument works against both.

to the supposedly felt aspects of these experiences, and say *more aspects of mental life of this kind.* A set of examples, and how they appear relevantly similar to us, are thus intended to fix the meaning of 'qualia'.[8] The realist and the eliminativist may agree on this identification strategy. They may agree on the set of examples and how we should take them to be relevantly similar – both may, for instance, have no difficulty generalising from existing cases to new examples. What the realist and eliminativist disagree about is whether the phenomenal feelings that *appear* to be present in these cases *really are* present. The realist claims that the experiences have, or instantiate, a property, 'what-it-is-like-ness', which should be added to our ontology. The 'what-it-is-like-ness' or phenomenal character is a real property of experience – as real as anything else. The realist says that explaining what this phenomenal property is, how it comes about, and how it relates to physical properties is the job of a theory of consciousness. In contrast, the entity eliminativist says that no such property exists (or is instantiated in the relevant cases). According to her, the examples are, in a sense, deceptive: they *appear* to show instantiation of a property, but that appearance is mistaken. There is no such property of experience.

An entity eliminativist denies the existence of qualia, but she does not deny that we have judgements, beliefs, and desires about qualia. This allows her to agree with a lot of what a realist says about experience. She can agree that we *believe* that our experience has qualia. She can agree that it is *hard for us to doubt* that our experience has qualia. She can agree that our beliefs and judgements about qualia *motivate us to act* according to their content. Nevertheless, the eliminativist says, these beliefs and judgements are false. Our beliefs and judgements about qualia are comparable to those of the ancient Greeks about Zeus: deeply held and capable of motivating action, but fundamentally mistaken. We should no more take on the project of explaining what qualia are, how they arise, and how they relate to physical properties than we should for Zeus.

We divide the rebooted version of Dennett's argument into two steps. The first step aims to establish an epistemic claim: we do not know which (diet) quale our experience instantiates. The claim here goes further than mere failure of infallibility or incorrigiblity: it is that we lack *any knowledge at all* of which quale our experience

---

[8]The specific examples cited in Chalmers (1996), Chapter 1 appear to play this role. Schwitzgebel (2016) outlines a similar strategy, although with the commitment that there should be a 'single obvious folk-psychological concept or category that matches the positive and negative examples'. (We do not think that either the realist or eliminativist need admit this.) Nida-Rümelin (2016), Section 3 outlines a similar strategy to identify 'experiential' properties, although she argues that if this strategy works, there can be no possibility of failure to refer so eliminativism is precluded. The general strategy of reference fixing by ostending examples from a single kind follows roughly the model used by a causal theory of the reference, although in the case of qualia the subject's ostensive relation to examples need not be causal (e.g. it could be some sort of non-causal acquaintance relation).

instantiates. The second step leverages this qualia scepticism to argue for qualia eliminativism. If the instantiation of one quale rather than another is unknowable, even in principle, then instantiation of qualia is a difference that makes no difference to the world; on that basis, qualia should be eliminated.

As a starting point, notice that sometimes it is hard to tell which quale your experience instantiates. Slow or subtle shifts in experience may leave you uncertain about which subjective feeling you have – is the feeling you have now on looking at an Yves Klein blue painting the same as that you had a minute ago? Such cases may present us with epistemically 'bad' cases of qualia knowledge: situations where for some reason we are unsure about which quale we have. Showing that there are 'bad' cases, however, does not show that we can never know which quale we have. It cannot sustain an eliminativist argument of the form just described.[9]

Focus instead on supposedly 'good' cases of qualia knowledge: cases where we appear to know, with a high degree of certainty, whether our qualia have changed across two conditions. 'Good' cases generally involve sudden or dramatic changes in experience. If the blue Yves Klein painting in front of you were suddenly swapped with a yellow painting, you would know, not only that the painting had changed, but also that your accompanying *phenomenal experience* had changed (perhaps you would regard the latter as part of your evidence base for the former). Dramatic changes in colour experience appear to provide a particularly good case of qualia knowledge. However, even in such 'good' cases, there are reasons to think that one lacks knowledge about which quale one has.

Suppose that while Lara is asleep a neurosurgeon operates on her brain. On waking, she finds that the world 'looks different' to her: objects that looked blue now look yellow. No one else notices the change, so Lara concludes that something must have happened to her brain. *Prima facie*, Lara appears to have reason to think that her qualia have changed. After all, 'things look different' now: she has qualitatively different experiences. However, the change she notices between the two conditions is compatible with two hypotheses:

(Q) Lara's qualia have changed from those she had yesterday
(R) Lara's qualia remain the same but her memories of her past experiences have changed

Lara's post-surgery experiences would be introspectively indistinguishable under either hypothesis Q or R. On the basis of introspection, Lara knows that 'things

---

[9]It is unfortunate that many of Dennett's intuition pumps involve subtle and slow changes in qualia. This has focused attention on failures of infallibility and incorrigibility in 'bad cases'. The more worrisome lesson is that there is no knowledge of qualia identity even in supposedly 'good' cases.

look different' now, but she cannot tell what causes them to look different: a change in her memory, a change in her current qualia, or some combination of both. Her problem is more serious than a lack of certainty or failure of infallibility. Q and R appear to be *equally supported* by Lara's introspective evidence. Introspection thus gives her *no knowledge at all* about whether her qualia have changed. As far as her introspective evidence base goes, she has no reason at all to favour one hypothesis over the other.

Lara, however, has access to other sources of knowledge than introspection. What if she were to look at changes in her brain across the two conditions? For the sake of argument, put Lara into the strongest possible epistemic position with regard to the physical state of her brain. She has perfect neuroscientific knowledge and full scans of her brain before and after surgery. Furthermore, suppose (unrealistically, but helpfully for Lara) that there is a separation in the neural basis of Lara's sensory systems and the neural basis of her memory systems. She can then reason as follows:

> If the brain scans reveal that the neural change affected only my sensory system and left my memory system intact, I have reason to favour Q over R because only the neural systems that support my current experience, and not my memories of past experiences, have been affected. Conversely, if the scans reveal that the neural changes affected only my memory system and left my sensory system intact, then I have reason to favour R over Q because only the neural systems that support my memories, and not those that support my current experience, have been affected.

Thus it seems that empirical evidence can do for Lara what introspection alone cannot: it can give her reason to favour Q or R. (Of course, it is possible that both Lara's sensory and memory systems are affected by the surgery, but we will ignore this as it would not help her.)

Unfortunately, Lara's reasoning depends on an assumption she has no reason to believe is true. Brain scans provide her with information about changes to her *neural events* but this can only help her with Q and R if she knows *where* in the causal order of those neural events her qualia experience occurs. Lara, however, does not know this, and by hypothesis it is not part of the concept of diet qualia. Consider two hypotheses about where her qualia experience might occur in the causal order of neural events:

1. Sensation → *qualia experience* → memory access
2. Sensation → memory access → *qualia experience*

Lara's reasoning assumed something like (1) is true: her qualia experience occurs *after* her current sensation but *before* memory access. Her memory is causally

downstream from her qualia experience; her qualia experience 'screens off' her sensations from her memory. This means that any surgery-induced changes to her memory system should have a differential effect on her qualia experience to changes to her sensory system: one affects something after that qualia experience, the other affects something before. If Lara discovers a neural change exclusively to her memory system, that suggests her current qualia experience is unaffected, because the causal antecedents of that experience remain the same as before. However, if she discovers a neural change to her sensory system (e.g. swapping her sensory channels) and no change to her memory, that suggests that her current qualia experience *has* been affected. We are in the realm of causal inference here – which falls short of deductive certainty (a cause could conceivably produce any effect). But this reasoning gives Lara *some* reason to prefer Q or R.

If hypothesis (2) is true, then Lara's inference above is invalid. Her qualia experience is causally downstream from both her sensation and her memory. Surgery-induced changes to the neural basis of *either* system could potentially affect her current qualia experience. A change to the neural basis of her memory could produce a change in her current qualia *or* a change in her memory of past experiences, or both. A change to the neural basis of her sensory system could produce a change in her current qualia *or* a change in the outputs of her memory systems, or both. The two aspects of her brain about which Lara can detect change (the neural bases of her memory and sensation) both lie causally upstream from her qualia experience, and so are confounded in the causal inference.

The problem is that no one knows whether (1), (2), or any number of other proposals about the order of qualia experience in neural processing is correct. This is not a limit of technology – a better scanner or more neural information would not help. Nor is it something with which the concept of diet qualia can help: the concept of diet qualia is silent about the causal role qualia play in perception and cognition. One might attempt to correlate Lara's introspective reports of qualia experience with her neural events to find out where among those events her qualia experience falls. However, we have already seen that there is no reason to trust Lara's introspective reports about the onset of or changes to her qualia in this context.[10] Neither introspection, nor empirical knowledge, nor some combination of the two tells Lara which quale she has. Even in an apparently 'good' case, with what looks like a clear contrast between the two conditions, Lara has no reason at all to favour Q or R. Given our own epistemic position is weaker than that of Lara, for all we know our own qualia may be changing all the time without us being able to detect it.[11]

---

[10]Other problems with such efforts are described in Section 5.2.

[11]Dennett (2005), Chapter 4 presents a similar argument for eliminating qualia using the phenomenon of change blindness, which again relies on cross-time comparisons. (For all we know, our colour qualia may be changing all the time without us noticing.)

This accounts for Step 1 of the argument. Step 2 says that if the identity of one's current quale is in principle unknowable (either by introspection or by methods available to empirical science), then we should eliminate qualia from our ontology. The reasoning is that qualia are an extra 'wheel' that do not turn anything in our ontology. Qualia have no characteristic, discernible effects – for if they did, Lara could use those effects to detect them. A quale's effect on us is always confounded with that of other factors (such as memory). Therefore, affirming the existence of qualia as independent, free standing entities in our ontology seems unmotivated. Like Wittgenstein's 'beetle',

> [This] thing in the box has no place in the language-game at all; not even as a something; for the box might even be empty – No, one can 'divide through' by the thing in the box; it cancels out, whatever it is. (Wittgenstein, 1958, §293)

It is open for a realist to *insist* that qualia should be included in our ontology irrespective of our inability to detect them (similarly, a beetle realist could stick to their guns and say that there really is a beetle in the box). But this seems at best unmotivated and at worse an expression of a dogmatic commitment to realism about qualia no matter what. In light of the problems above, qualia do not appear to earn their ontological keep. Better to say that, as inessential cruft, they should be eliminated.

Unlike Dennett's original argument, this rebooted argument does not rely on the assumption that qualia are intrinsic, private, ineffable, or directly accessible. Before closing, we with to flag two problems with this new argument.

First, one might wonder, even if the argument is correct, why it still *seems* to us that qualia exist. The 'seeming' does not go away even if one embraces the eliminativist's conclusion. On this basis, one might press for a residual role for qualia that provides more than an eliminativist would allow: more than merely being associated with a set of dispositions to make judgements, or with having beliefs about qualia (both compatible with those judgements and beliefs being false). Our relationship to qualia appears to be more primitive than this. It *seems* to us that our experiences have qualia and this 'seeming' is the evidence for our beliefs about qualia. But how can this impression, this pre-doxastic 'seeming', be produced? One might tell a mechanistic and adaptationist story about how humans arrive at their false beliefs about qualia (Dennett, 1991; N. Humphrey, 1992). But what mechanistic story can be told that explains the production of *seemings* that generate and appear to confirm these beliefs? This is the 'illusion problem', discussed in the next section and it remains an unsolved challenge for qualia eliminativists.

Second, one might question whether the qualia scepticism of Step 1 is sufficiently

motivated. Step 1 relies on questioning the reliability of memory-based comparisons: Does Lara know whether a quale she has today is the same as the one she had yesterday? A qualia realist might concede that she does not know this (perhaps because of confounds with memory), but deny that Lara lacks *any knowledge at all* of which quale her experience instantiates. Imagine looking out on a mountain scene with green grass, grey rock, and blue sky. Multiple qualia are instantiated simultaneously in your experience: *what-it-is-like to see green*, *what-it-is-like to see grey*, *what-it-is-like to see blue*, and so on. You can tell the difference between these qualia (you can make similarity judgements, detect that there are many qualia versus a few, distinguish between your visual, auditory, and proprioceptive qualia, and so on). None of these judgements appear to rely on memory. Within the domain of current experience, therefore, you appear to have *some* knowledge about which qualia your experience instantiates. But then, why think that qualia are a wheel that turns nothing or are always unknowable and confounded with memory in their effect on you?

4.3   The illusionist movement

Recently, interest in eliminativist approaches to phenomenal aspects of consciousness has been rekindled by Frankish, in particular in a position paper and a series of responses in a special issue of the *Journal of Consciousness Studies*. Frankish outlines 'illusionism' as the view that experiences have no phenomenal properties and that our phenomenal feelings are 'illusory'. We think we have experiences with phenomenal properties, but in fact we do not. Illusionism is a form of entity eliminativism about phenomenal consciousness even if the label 'eliminativism' is avoided for rhetorical reasons. It is motivated somewhat differently to Dennett's entity eliminativism, and has a slightly different focus, so is worth discussion in its own right.[12]

First, illusionism is partly motivated by taking seriously the idea that phenomenal properties, and phenomenal consciousness, cannot be accounted for scientifically. Illusionism is seen as a way out of this 'explanatory gap' problem. Second (and relatedly), the reasons for favouring illusionism are mainly rather general, theoretical reasons. The theoretical virtue of simplicity, or conservativism, suggests that the

---

[12]Dennett-style eliminativism treats our ontological commitment to phenomenal consciousness as a theoretical mistake: there is nothing that satisfies the description of qualia, or qualia are ontologically inert and therefore it is safe to eliminate them. Somewhat differently, one can see illusionism as treating our ontological commitment to phenomenal consciousness as an introspective or perceptual mistake: we 'perceive' (via introspection) that our experience has phenomenal properties but it does not (hence the illusionism title). However, see Frankish (2016b) for ways of blurring the boundary between theoretical and introspective/perceptual mistakes.

fewer entities/properties the better. Since illusionism gets rid of the metaphysically and epistemically problematic phenomenal properties, illusionism is better than alternative realist positions. Third, illusionism is often argued to be a research programme rather than a set of worked-out claims, and this research programme is worth pursuing more than its alternatives. As we will see, illusionism comes with a range of difficult open questions.

Illusionism follows a slightly different tack to the typical argument for entity eliminativism described in Section 2. The first step is supposed to be identifying the contested entity/property in a way that can be generally accepted. This is not straightforward for phenomenal properties (see discussion both above and below). Second, the arguments motivating illusionism are not direct arguments to the effect that phenomenal properties, as described, do not exist; the position is largely motivated on other grounds (e.g. theoretical simplicity). The third step of the classic argument is to conclude that phenomenal properties do not exist. This is also concluded by some proponents of illusionism, but one could arguably also treat illusionism as a promising research programme without in advance committing to this conclusion.

Challenges to illusionism come in roughly three forms (the first two roughly track two of the steps above).

First, one might argue that it is neither obvious nor universally accepted what phenomenal consciousness is, or what phenomenal properties are, such that a proposal to eliminate them is comprehensible. Mandik (2016) states that 'phenomenal' is a technical (not folk) term, but one that is not clearly defined. As such, both eliminativist and realist talk about 'phenomenality' is unwarranted; in neither case is there a clear target to be eliminativist or realist about. Schwitzgebel (2016) tries to provide a minimal 'definition by example' that is not committed to any particular (troublesome) metaphysical or epistemic commitments, but as Frankish (2016b) points out, this is not substantive enough to sway the debate one way or the other.

Second, one might reject some of the main theoretical motivations for thinking illusionism is the best or most reasonable philosophical position available. For example, Balog (2016) defends the phenomenal concept strategy, which concedes the existence of an insuperable explanatory gap but still tries to preserve realism about phenomenal properties. Prinz (2016) also defends a realist account of phenomenal properties, but one that tries to close the explanatory gap by providing neuroscientific explanation of at least some aspects of phenomenal consciousness. More generally, unless one is convinced that the theoretical virtues of illusionism (parsimony, fit with existing science) are superior to rival positions on consciousness, one is unlikely to persuaded of illusionism.

Third, a cluster of worries arise around the 'illusion problem'. The illusion problem

is how to account for the alleged illusion of phenomenality. How can one have experiences that appear to have phenomenal properties without any phenomenal properties existing? Frankish (2016a) labels those physical properties (perhaps highly disjunctive and gerrymandered) that typically cause our introspection to misrepresent us as having phenomenal qualities, 'quasi-phenomenal properties'. Quasi-phenomenal redness is, for example, the physical property that typically causes (false) representations of phenomenal redness in our introspective experience. According to Frankish, it is the tokening of these false introspective representations that is responsible for the illusion of phenomenal consciousness. He likens the effect to that of other persistent, false perceptual representations such as our perception of impossible figures like the Penrose triangle (N. Humphrey, 2011) or of colours as 'out there' in the world (Pereboom, 2011).

The worry is how exactly this is supposed to work. It is not clear how a false representation caused by non-phenomenal properties could produce an appearance or 'seeming' of phenomenality. And as Prinz puts it, '… what is it about beliefs in experience that causes an illusion of experience?' (2016, p. 194): how and why is it that these representations cause illusions of subjective experience when other sorts of false representations do not? Related to this is a worry about how such false introspective representations get their content (Balog, 2016). Representations of phenomenal feelings are not like other false or non-referring representations ('Pegasus', 'the largest prime'). These appear to get their content by being semantic constructs from representations that do refer ('winged', 'horse', 'largest', 'prime'). How do representations of phenomenal experience get their content if there are no phenomenal feels, and if representations of phenomenal feels are not semantic constructs built from representations that do refer?

Illusionism promises to get us away from the hard problem. It simply eliminates the 'data' the hard problem asks us to physically explain – phenomenal feelings. Prinz (2016) argues that the illusion problem and the hard problem in fact face similar difficulties. In both cases, we need to identify what phenomenal properties are. In the hard problem, we need to explain how phenomenal properties come out of 'mere matter': how feelings arise in an apparently non-phenomenal system. In the illusion problem, we need to explain how (vivid!) illusions of phenomenality come about in entirely non-phenomenal systems. The challenge is to explain how an illusion of phenomenality (worthy of that name) arises in a non-phenomenal system. In both cases, then, one needs to explain how something suitably like phenomenality arises from 'mere matter'. By the time one has done this, it might be just as easy to be a realist as an illusionist.

Frankish (2016a) briefly considers the relationship between illusionism and discourse eliminativism: 'Do illusionists then recommend eliminating talk of phenom-

enal properties and phenomenal consciousness? Not necessarily' (p. 21). He goes on to suggest that discourse eliminativism could be avoided by an illusionist if the phenomenal terms in science were redefined to refer to quasi-phenomenal properties – the physical properties that typically cause the relevant false introspective representations. This seems to us neither necessary nor likely.

First, as Frankish says, it would be a substantial departure from what these terms mean in other contexts, and so it would invite confusion. Second, although we agree that it is likely that an illusionist scientific psychology would need a way to keep track of quasi-phenomenal properties, this could most naturally be done with a response-dependent characterisation of those properties: group together the physical properties that give rise to a specific (false) phenomenal representation and type quasi-phenomenal properties accordingly. Keeping track of those quasi-phenomenal properties does not require redefining the language of phenomenal terms in science. Third, as Frankish (2016b) says, it is no part of illusionism to say that the illusion of conscious experience is not important or useful to the experiencer. Graziano (2016) and Dennett (1991) argue that phenomenal consciousness plays a central, adaptive role in our mental lives. It is reasonable to expect that a scientific psychology would want to study this role. That could be done while bracketing questions about the existence of phenomenal properties.[13] Similarly, a psychology that studied childhood dreams might want to talk about the role of representations of Santa Claus and unicorns – without attempting to redefine those terms to refer to physical entities, or positing real entities corresponding to those terms. Talk of phenomenal feels will likely remain in scientific psychology, albeit with the qualification that the entities that appear to stand behind this talk do not exist.

5   Discourse eliminativism about consciousness

We now turn to discourse eliminativism. Discourse eliminativism seeks to rid science of talk, concepts, and practices associated with phenomenal consciousness (even if phenomenal consciousness is still admitted to exist). In this section, we look at three discourse eliminativist arguments. The first is based on considerations raised by psychologists at the start of the twentieth century. The second is based on more contemporary concerns about how to study phenomenal consciousness independently of access consciousness and the mechanisms of reportability. The third is based on the worry that the concept of consciousness fails to pick out a scientifically usable category of phenomena.

---

[13]Dennett's (1991) heterophenomenology provides one model of how an illusionist might do this.

5.1   Scientific behaviourism

One of the goals of scientific psychology in the first half of the twentieth century was to redefine psychology, not as the study of the mind, but as the study of observable behaviour. Scientific behaviourists argued that scientific psychology should avoid talk of internal mental states, and in particular talk of conscious states (Hull, 1943; Skinner, 1953; Watson, 1913).

The rise of scientific behaviourism was at least partly due to the perceived failure of an earlier attempt to scientise psychology via trained use of introspection (Titchener, 1899). The debate on the nature of imageless thought was held up as an example of how unproductive that research programme was. One side in the debate appealed to introspection to argue that all thoughts were analysable into images; the other side used similar evidence to argue that some thoughts are imageless. The disagreement was widely seen as unresolvable because the evidence from disagreeing enquirers could not be compared in an unbiased way. By the mid-twentieth century, introspective methods were discredited and scientific study of conscious experience has been largely abandoned (G. Humphrey, 1951).

Scientific behaviourists sought to reposition psychology to avoid these methodological difficulties. The subject matter of scientific psychology should be publicly observable, publicly verifiable, or independently experimentally controllable events. Psychology should eliminate talk of conscious experience and use of introspective methods. However, this did not mean that behaviourists thought that internal mental states, including states of consciousness, did not exist: 'The objection to inner states is not that they do not exist, but that they are not relevant in a functional analysis' (Skinner, 1953, p. 35).[14] Scientific behaviourists proposed an alternative way of talking, thinking, and acting that they argued was superior (in predictive, explanatory, and methodological terms) to a scientific psychology that appealed to, or attempted to study, conscious experience. Phenomenal consciousness, irrespective of its ontological reality, should be excluded from scientific psychology because its study was methodologically flawed and appeal to it was unnecessary to explain behaviour.

Based on related considerations about verification and public accessibility, positivistically inclined philosophers argued that various ontological and/or semantic lessons follow with regard to conscious experience (Ryle, 1949; and less clearly, Wittgenstein, 1958). This prompted them to redefine mental state language in terms of behavioural dispositions and/or to eliminate qualitative conscious feelings from ontology. However, to connect these two lines of thought – one about scientific practice and the other about ontology/semantics of ordinary language – requires

---

[14]See Hatfield (2003) for discussion of the views of other behaviourists.

accepting auxiliary claims about verifiability, the role of science, and the scope of human knowledge. Such links are widely questioned today. Many scientific behaviourists did not perceive such links either and they argued for the elimination of talk of conscious experience from science based on pragmatic rather than meta-ontological concerns.

5.2 Eliminativism via independent access

A different methodologically motivated form of discourse eliminativism about phenomenal consciousness is found among some consciousness researchers today. This stems from the problems involved in trying to operationalise consciousness, or finding ways to experimentally probe it. One typical way of operationalising consciousness is via some kind of reportability: a subject is conscious of a stimulus if and only if they report it or respond to it in some way. This is fairly straightforward, but there are potential problems in using reportability as a marker of the presence of phenomenal consciousness, rather than as a marker of the cognitive capacities associated with consciousness. This can motivate a position of discourse eliminativism about phenomenal consciousness.

First, consider the distinction between phenomenal and access consciousness. Phenomenal consciousness refers to felt conscious experiences, (diet) qualia, raw feels, and so on. Access consciousness refers to the aspects of consciousness that are associated with or can be used in cognitive capacities like reasoning, action, verbal report, and so on. If we somehow knew that access and phenomenal consciousness were always bound together (no cognitive access without phenomenal consciousness and vice versa), then scientific ways of probing access consciousness would also function as scientific ways of probing phenomenal consciousness. That is, if phenomenal consciousness and aspects of consciousness related to cognitive access always go together, then probing these cognitive capacities just is to probe phenomenal consciousness. In this case (absent any other problems), it would be perfectly legitimate for the term 'phenomenal consciousness' to figure in scientific discourse, because the phenomenon it picks out is scientifically accessible.

The problem is that it is not obvious whether the aspects of consciousness picked out by access and phenomenal consciousness are always co-present. According to Block (1995), there may be instantiations of phenomenal consciousness (raw feels) without any related cognitive access (ability to respond to or report about these raw feels). Block has outlined a number of examples where this might happen, including when subjects may have highly detailed and specific phenomenal experiences, but be unable to report the details of them (Sperling paradigm); cases of phenomenal consciousness of unattended items; and possibly cases of hemi-spatial neglect, where

subjects do not appear to have phenomenal experiences from some part of their visual field (see Block, 2007; Block, 2011; Block, 2014; Irvine, 2011; Phillips, 2011, for discussion of some of these cases). In most of these cases, there is evidence that subjects are at least processing sensory information that they are unable to report about. Block's claim is that there is a layer of untapped and unaccessed phenomenal consciousness present in these cases, in addition to whatever can be overtly reported or measured.

The lack of a way to probe the phenomenal aspect of consciousness independently of the accessibility aspect makes it difficult (or impossible) to scientifically assess these claims. It looks like any way of probing phenomenal consciousness requires that the experience had some measurable effect on the subject, possibly such that she can report it in some way. That is, accessing phenomenal consciousness relies on it being associated with some kind of cognitive function or capacity, therefore accessing phenomenal consciousness relies on it being associated with access consciousness. So, if an instance of phenomenal consciousness is not associated with access consciousness, then it looks like we cannot tell if it is present or not. As Dehaene et al. (2006) note, whether participants in an experimental situation 'actually had a conscious phenomenal experience but no possibility of reporting it, does not seem to be, at this stage, a scientifically addressable question' (p. 209).

Partly in response to this worry, Block, Lamme, and colleagues have argued for the possibility of *indirectly* investigating these purported instances of phenomenal consciousness without accessibility (Block, 2011; Block, 2014; Lamme, 2006; Sligte et al., 2010). The idea here is to find some reasonable and measurable marker of the presence of consciousness in cases where phenomenal (and access) consciousness is clearly present (call this marker M). This could be a particular neurophysiological signature (e.g. evidence of strong feed-forward processing), or a particular behavioural marker (e.g. ability to complete a particular type of task based on a set of visual stimuli). One then argues that if the special marker M is present in a subject, then regardless of whether the subject appears to be conscious of the test stimulus according to other standard measures of (access) consciousness, the subject is phenomenally conscious of the test stimulus. That is, special marker M guarantees that a subject is phenomenally conscious of the test stimulus, even if they don't report seeing it, or can't perform a range of actions that we usually associate with being conscious of a stimulus. The subject experiences phenomenal consciousness of the stimulus without having cognitive access to that experience.

However, problems of interpretation abound here. Such behavioural and neurophysiological evidence could be taken as indirect evidence of phenomenal consciousness without access consciousness, but it could also be interpreted as evidence of unconscious processing (if we got the special marker M wrong), or of graded

cognitive access and phenomenal consciousness of the stimulus (see replies to Block, 2007). There are no direct scientific grounds on which to choose between these interpretations, because there is no direct way to assess whether marker M has anything to do with phenomenal consciousness.

One response to these discussions is to advocate discourse eliminativism about phenomenal consciousness. This is based on accepting that there is no direct way to probe phenomenal consciousness independently of cognitive access, and that there are no straightforward empirical ways of testing the claim that phenomenal consciousness can be present independently of cognitive access. In this case, the only aspect of consciousness that can definitely be probed scientifically is cognitive access, that is, access consciousness. In terms of achievable scientific practice, the safest methodological route is to drop talk of phenomenal consciousness. Something like this position appears to be taken by a number of consciousness researchers (possibly including Dehaene).

This position is compatible with a range of positions about the ontology of phenomenal consciousness. One might be either an entity eliminativist or an entity realist about phenomenal consciousness. One might claim that phenomenal consciousness can (possibly or probably) exist without cognitive access, or be agnostic about this possibility. Alternatively, one might argue, with Cohen and Dennett (2011), that if a phenomenally conscious state is not accessible to scientists or the subject having it (e.g. via some kind of report), then it is (evolutionarily, cognitively) implausible to call it a state of consciousness at all. In this case, if phenomenal consciousness exists, it always co-occurs with cognitive access.

5.3   Eliminativism via identity crisis

The kind of discourse eliminativism about phenomenal consciousness outlined above is based on a problem about *accessing* the phenomenon in question. Another kind of discourse eliminativism about consciousness is based on the problem of *identifying* the phenomenon in question. For now, assume that phenomenal consciousness always co-occurs with access consciousness (perhaps for the reasons suggested by Cohen and Dennett above), so that we can (for the minute) just work with the term 'consciousness' which will pick out both. However, even with this problem out of the way, it is still questionable whether the concept of consciousness picks out a clear category of phenomena that is scientifically useful. If it does not, this provides a new motivation for discourse eliminativism about consciousness, and (by assumption) discourse eliminativism about phenomenal consciousness too.

Above, it was suggested that there is a reasonably broad consensus that assessing

the presence or absence of consciousness has something to do with reportability. However, reportability can be realised in a number of ways, some of which are incompatible with each other (see Irvine, 2013, for review). One 'objective measure' of consciousness from psychophysics relies on forced-choice tasks, where subjects are typically shown a masked stimulus for a very short period of time and are 'forced' to choose between two response options (e.g. stimulus present/absent, stimulus was a square/circle). On the basis of these responses, the subjects' underlying 'sensitivity' to the stimuli is calculated. The resulting objective measure is highly stable and not subject to problematic biases, but it is liberal, and often attributes visual consciousness of stimuli to subjects who explicitly deny having any. As a result, it is sometimes criticised as being merely a measure of sensory information processing (e.g. Lau, 2008). Despite being acknowledged as problematic, objective measures such as these tend to be used in studies of visual consciousness because of their desirable properties as scientific measures (they are stable, bias-free).

In contrast, 'subjective measures' of consciousness tend to use free reports or a wide range of possible responses, sometimes generated in advance by the subjects themselves. The experimental methodology may be based around emphasising to subjects careful use of introspection, assessing subject's confidence in their reports (sometimes using wagering), or just recording simple, untutored responses. Subjective measures get much closer to what subjects themselves acknowledge about their conscious experience. However, the precise ways that subjective measures are generated can have a significant impact on whether consciousness is deemed to be present or absent (or somewhere in between) (Sandberg et al., 2010; Timmermans and Cleeremans, 2015). As scientific measures, they are highly unstable and subject to bias. They also regularly conflict with objective measures of consciousness (except under artificial training conditions), and are generally accepted as being conservative (they do not normally capture all instances of conscious experience).

The difficulties in identifying an adequate measure of consciousness reappear in debates about the neural correlate(s) or mechanism(s) of consciousness. The difficulty is that behavioural measures are key in identifying these correlates and mechanisms. Roughly speaking, one chooses a measure; identifies the neural activity that occurs when the measure says that consciousness is present; and treats this as 'the' mechanism of consciousness. However, using different behavioural measures (unsurprisingly) leads to the identification of different neural correlates. These span from 'early' neural activity for some liberal behavioural measures (which may capture early sensory processing), to 'late' and attention-based neural activity for conservative behavioural measures (which may capture later cognitive uptake of the conscious experience) (see Irvine, 2013). With no settled agreement on what counts as an adequate measure of consciousness, there can be no agreement on what the neural correlates and mechanisms of consciousness are.

This plethora of measures and mechanisms is not necessarily problematic in itself, but Irvine (2012) argues that there is no methodologically viable way of resolving problems about how to adjudicate between them when they conflict. Each measure has its pros and cons, but none exists that is both scientifically adequate (i.e. fairly stable over repeated measures and bias-free) and fits with pre-theoretic commitments about what consciousness is. To choose one measure would be to (operationally) define consciousness by fiat, which would undermine the motivations for engaging in 'real' consciousness science in the first place. Furthermore, the mechanisms that correlate with these varied measures do not form a well-demarcated scientific kind, or even a well-demarcated group of kinds: they have no more in common than any arbitrary group of mechanisms of perception and cognition. They range over mechanisms of sensory processing, attentional processes, and processes related to decision making, report, and meta-cognition.

This provides one set of reasons for eliminating talk of consciousness in serious science. There are a wide range of incompatible things that 'consciousness' could pick out, and no methodologically acceptable way of deciding between them. If a scientific concept is surrounded by serious methodological problems, then (if they are bad enough) this is motivation for eliminating the concept from science. This consideration is reinforced by pragmatic concerns in favour of eliminating talk of consciousness. Given that it is unclear what 'consciousness' refers to, this generates unproductive debates and miscommunication; it blocks the generation of useful predictions and generalisations; and it promotes misapplications of research methodologies and heuristics. That is, there are negative consequences from continued use of the concept 'consciousness' in scientific practice.

There is also a better alternative. This alternative demands that researchers use terms that clearly demarcate the phenomena under study, potentially by referring to how they are experimentally operationalised. This could done by splitting up the phenomena by how they are measured (e.g. forced-choice tasks, confidence ratings, or free report). Using these more specific terms avoids the problems above. By precisely specifying what the phenomenon is and how it is measured, there is no ambiguity about what phenomenon is picked out. This also makes it possible to identify the mechanism that generates it, make robust predictions and generalisations about the target phenomenon, and avoid miscommunication.

As before, discourse eliminativism is not tied to entity eliminativism. (Irvine's (2012) position does not entail entity eliminativism of any sort.) Discourse eliminativism is solely about scientific representations, concepts, methods, and practices, and what it is scientifically appropriate and useful to take as a target of study. Whatever 'consciousness' (access or phenomenal) refers to may be out there, even if it is not a useful scientific concept.

6 Conclusion

In this chapter, we have reviewed a range of arguments for positions of entity and discourse eliminativism, primarily aimed at phenomenal consciousness. Entity eliminativists deny the existence of phenomenal consciousness, while discourse eliminativists deny the scientific utility of talking about phenomenal (and perhaps access) consciousness.

Entity eliminativism can be pursued in a number of ways. One standard method is to describe the entity in question, and then show that nothing exists that satisfies that definition (4.1). This can be expanded to the method of using examples to fix the subject matter, and generalising to the relevant entity or property (4.2). A third approach taken by illusionists (4.3) is to use a loose definition of the relevant entity/property, but to argue that whatever this refers to, it is theoretically and metaphysically simpler and more productive to work on the assumption that it does not exist. One problem that faces entity eliminativists of all types is to solve the 'illusion problem', which is the parallel of the hard problem faced by realists about phenomenal consciousness. This is the problem of how something non-phenomenal can give rise to something that seems phenomenal, or fools us into thinking that there are phenomenal properties.

Positions of discourse eliminativism are also motivated in slightly different ways, though all focus on scientific methodology. Classic behaviourism focused on what could be unequivocally measured in a public and 'observable' way, eradicating talk of mental states (5.1). More recent scientific work on consciousness has tended to move away from discussion of phenomenal consciousness on the basis that it is not clear whether scientific methodology can probe it independently of the cognitive abilities associated with access consciousness (5.2). An argument can be made that the general concept of consciousness should be eliminated from scientific talk given the problems in clearly demarcating the phenomenon in question (5.3). Eliminating discourse about phenomenal consciousness from science might seem to remove a key concept in explaining human behaviour. However, this is not necessarily the case: specific reports and judgements about phenomenal consciousness can still function in explanations, and as explanatory targets in their own right.

Acknowledgements

Bibliography

Balog, K. (2016). "Illusionism's discontent". In: *Journal of Consciousness Studies* 23, pp. 40–51.

Block, N. (1990). "Consciousness and accessibility". In: *Behavioral and Brain Sciences* 13, pp. 596–598.

— (1995). "On a confusion about a function of consciousness". In: *Behavioral and Brain Sciences* 18, pp. 227–247.

— (2007). "Consciousness, accessibility, and the mesh between psychology and neuroscience". In: *Behavioral and Brain Sciences* 30, pp. 481–548.

— (2011). "Perceptual consciousness overflows cognitive access". In: *Trends in Cognitive Sciences* 15, pp. 567–575.

— (2014). "Rich conscious perception outside focal attention". In: *Trends in Cognitive Sciences* 18, pp. 445–447.

Carruthers, P. (2000). *Phenomenal Consciousness*. Cambridge: Cambridge University Press.

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

Cohen, M. A. and D. C. Dennett (2011). "Consciousness cannot be separated from function". In: *Trends in Cognitive Sciences* 15, pp. 358–364.

Dehaene, S., J.-P. Changeux, L. Naccache, J. Sackur and C. Sergent (2006). "Conscious, preconscious, and subliminal processing: A testable taxonomy". In: *Trends in Cognitive Sciences* 10, pp. 204–211.

Dennett, D. C. (1988). "Quining qualia". In: *Consciousness in Contemporary Science*. Ed. by A. J. Marcel and E. Bisiach. Oxford: Oxford University Press, pp. 42–77.

— (1991). *Consciousness Explained*. Boston, MA: Little, Brown & Company.

— (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press.

Frances, B. (2008). "Live skeptical hypotheses". In: *The Oxford Handbook of Skepticism*. Ed. by J. Greco. Oxford: Oxford University Press, pp. 225–244.

Frankish, K. (2012). "Quining diet qualia". In: *Consciousness and Cognition* 21, pp. 667–676.

— (2016a). "Illusionism as a theory of consciousness". In: *Journal of Consciousness Studies* 23, pp. 11–39.

Frankish, K. (2016b). "Not disillusioned: Reply to commentators". In: *Journal of Consciousness Studies* 23, pp. 256–289.

Graziano, M. S. A. (2016). "Consciousness engineered". In: *Journal of Consciousness Studies* 23, pp. 98–115.

Hatfield, G. (2003). "Behaviourism and psychology". In: *Cambridge History of Philosophy, 1870–1945*. Ed. by T. Baldwin. Cambridge: Cambridge University Press, pp. 640–648.

Hull, C. L. (1943). *Principles of Behavior*. New York, NY: Appleton-Century.

Humphrey, G. (1951). *Thinking*. London: Methuen.

Humphrey, N. (1992). *A History of the Mind: Evolution and the Birth of Consciousness*. New York, NY: Simon and Schuster.

— (2011). *Soul Dust: The Magic of Consciousness*. Princeton, NJ: Princeton University Press.

Irvine, E. (2011). "Rich experience and sensory memory". In: *Philosophical Psychology* 24, pp. 159–176.

— (2012). *Consciousness as a Scientific Concept: A Philosophy of Science Perspective*. Dordrecht: Springer.

— (2013). "Measures of consciousness". In: *Philosophy Compass* 8, pp. 285–297.

Kind, A. (2001). "Qualia realism". In: *Philosophical Studies* 104, pp. 143–162.

Lamme, V. A. (2006). "Towards a true neural stance on consciousness". In: *Trends in Cognitive Sciences* 10, pp. 494–501.

Lau, H. (2008). "Are we studying consciousness yet?" In: *Frontiers of Consciousness*. Ed. by L. Weiskrantz and M. Davies. Oxford: Oxford University Press, pp. 245–258.

Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.

Mach, E. (1911). *The History and Root of the Principle of Conservation of Energy*. Chicago, IL: Open Court.

Mallon, R., E. Machery, S. Nichols and S. P. Stich (2009). "Against arguments from reference". In: *Philosophy and Phenomenological Research* 79, pp. 332–356.

Mandik, P. (2016). "Meta-illusionism and qualia quietism". In: *Journal of Consciousness Studies* 23, pp. 140–148.

Nida-Rümelin, M. (2016). "The illusion of illusionism". In: *Journal of Consciousness Studies* 23, pp. 160–171.

Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*. Oxford: Oxford University Press.

Phillips, I. (2011). "Perception and iconic memory". In: *Mind and Language* 26, pp. 381–411.

Prinz, J. (2016). "Against illusionism". In: *Journal of Consciousness Studies* 23, pp. 186–196.

Quine, W. V. O. (1980). "On what there is". In: *From a Logical Point of View*. Cambridge, MA: Harvard University Press, pp. 1–19.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.

Sandberg, K., B. Timmermans, M. Overgaard and A. Cleeremans (2010). "Measuring consciousness: is one measure better than the other?" In: *Consciousness and Cognition* 19, pp. 1069–1078.

Schwitzgebel, E. (2016). "Phenomenal consciousness, defined and defended as innocently as I can manage". In: *Journal of Consciousness Studies* 23, pp. 224–235.

Searle, J. R. (1997). *The Mystery of Consciousness*. London: Granta Books.

Skinner, B. F. (1953). *Science and Human Behavior*. New York, NY: Macmillan.

Sligte, I. G., A. R. Vandenbroucke, H. S. Scholte and V. A. Lamme (2010). "Detailed sensory memory, sloppy working memory". In: *Frontiers in Psychology* 1, p. 175.

Sloman, A. and R. Chrisley (2003). "Virtual machines and consciousness". In: *Journal of Consciousness Studies* 10, pp. 113–172.

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

Timmermans, B. and A. Cleeremans (2015). "How can we measure awareness? An overview of current methods". In: *Behavioural Methods in Consciousness Research*. Ed. by M. Overgaard. Oxford: Oxford University Press, pp. 21–46.

Titchener, E. B. (1899). *A Primer of Psychology*. New York, NY: Macmillan.

Tye, M. (2002). "Visual qualia and visual content revisited". In: *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press, pp. 447–456.

Watson, J. (1913). "Psychology as a behaviorist views it". In: *Psychological Review* 20, pp. 158–177.

Wittgenstein, L. (1958). *Philosophical Investigations*. 2nd ed. Oxford: Blackwell.