

How experimental neuroscientists can fix the hard problem of consciousness

Colin Klein ^{1,*},[†] and Andrew B. Barron²

¹School of Philosophy, The Australian National University, Canberra, ACT 0200, Australia and; ²Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109, Australia

*Correspondence address. School of Philosophy, The Australian National University, Canberra, ACT 0200, Australia. Tel: +61 02 6125 2792; E-mail: colin.klein@anu.edu.au

[†]Colin Klein, <https://orcid.org/0000-0002-7406-4010>

Abstract

For the materialist, the hard problem is fundamentally an explanatory problem. Solving it requires explaining why the relationship between brain and experience is the way it is and not some other way. We use the tools of the interventionist theory of explanation to show how a systematic experimental project could help move beyond the hard problem. Key to this project is the development of second-order interventions and invariant generalizations. Such interventions played a crucial scientific role in untangling other scientific mysteries, and we suggest that the same will be true of consciousness. We further suggest that the capacity for safe and reliable *self*-intervention will play a key role in overcoming both the hard and meta-problems of consciousness. Finally, we evaluate current strategies for intervention, with an eye to how they might be improved.

Keywords: consciousness; intervention; hard problem; meta-problem

Introduction

The modern neuroscience of consciousness begins with a division of territory. In seminal work, the philosopher David Chalmers distinguished the easy and hard problems of consciousness (Chalmers 1998; Chalmers 2003). Easy problems involve sorting out the mechanisms that mediate conscious perception and action. The hard problem requires explaining why activity in these mechanisms is accompanied by any subjective feeling at all. Why, in Nagel's (1974) evocative phrase, is there something it is like for you to be you, while there's nothing it's like for a rock to be a rock? Or as Chalmers (1996, 201) puts it:

Why is it that when our cognitive systems engage in visual and auditory information-processing, we have visual or auditory experience: the quality of deep blue, the sensation of middle C? ... Why should physical processing give rise to a rich inner life at all?

Chalmers and others argue that the hard problem is a deep philosophical mystery, upon which empirical evidence could have little bearing.

Meanwhile, in several influential pieces Crick and Koch argued that, by focusing on easy problems, it might be possible to meaningfully work around the hard problem. They noted that 'at any one moment some active neuronal processes correlate with consciousness, while others do not' (Crick and Koch 1990, 263). Thus there is a viable scientific project that searches for the neural correlates of consciousness (NCCs). Chalmers did not disagree—his formulation of the hard problem is compatible with the existence of NCCs. So the search for NCCs could progress whether or not the hard problem has a solution.

Thus was the territory divided. Philosophers inherited the hard problem. Scientists got the search for NCCs. Despite occasional defectors on both sides, this truce has held for a quarter

Received: 1 August 2019; Revised: 30 March 2020. Accepted: 4 April 2020

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

century. Yet, the search for NCCs continues to hit an impasse that looks more philosophical than empirical (Hohwy and Bayne 2015). Conversely, philosophy divorced from neuroscience has endorsed a variety of counterintuitive views. At one extreme, there is defence of panpsychism, the position that consciousness is everywhere (Chalmers 2003). At the other, philosophers assert that consciousness is simply an illusion (Dennett 1991; Irvine 2013; Kammerer 2019). Such wildly differing views feel like stagnation, not progress.

Outside of the study of consciousness, meanwhile, both fields have made significant advances. Comparative neuroscience has made great strides investigating the evolutionary origins of the capacities that support consciousness (Feinberg and Mallatt 2016; Klein and Barron 2016; Ginsburg and Jablonka 2019). Philosophers of science have demonstrated the importance of interventions for explanation, while neurobiological technique for carrying out interventions has rapidly improved.

These developments have made possible a *rapprochement*. What follows outlines an experimental programme for making progress on the hard problem. The key argument is that the development of safe interventions on experience, including the capacity for self-intervention, is ultimately necessary for moving beyond the hard problem. This is a difficult task—but it is one that falls within the traditional bailiwick of experimental neuroscience. What we sketch is a proposal rather than a solution; the challenge is to show how experimentation might be useful, so as to guide the development and design of future projects.

Intervention, Explanation and Consciousness

We take the hard problem of consciousness to be a puzzle about scientific explanation (Levine 1983; Irvine 2013). The materialist must give a full explanation of *when* and *why* we are conscious that is grounded in materialistically respectable principles. Conversely, the property dualist bets that such explanations will require appeal to *sui generis* contingent laws that happen to hold here, but that could be absent.

Following many philosophers of science, we take a fundamentally *interventionist* approach to explanation. Interventionism has its roots in Pearl's (2000) seminal work on causation and has been developed most notably by Woodward (2000, 2003, 2010). The core idea is that we explain a phenomenon when we can demonstrate how aspects of it can be reliably changed by manipulating other parts of the world.

Interventionism has a direct connection to experimental practice. It has found traction in analysing the link between experiment and explanation, particularly in special sciences like genetics, neuroscience and economics. Though originally conceived as a way to explicate causal relationships, interventionism has also found important use in analysing synchronic relationships such as those between cognitive processes and the neural processes that realize them (Craver 2007; Woodward 2010; Klein 2017).

The core idea of the interventionist approach is the observation that changing the brain can reliably change what one experiences. Begin with the situation illustrated in Fig. 1a. Suppose we have discovered that a brain process B is reliably correlated with an experience P of redness. Now, this correlation might be an accident. It might also be due to a mere downstream effect of what's really doing the work. Regardless, it will hold only because there are a variety of background conditions and enabling conditions, the absence of which would remove consciousness altogether. Given that, what would it take to say that the

presence of B actually *explains*, on a particular occasion, why the subject experiences red?

A partial answer comes from realizing that there should be a way to intervene on B in order to change the experience of red to an experience of (say) blue. Figure 1b shows such a hypothetical intervention. The key idea is that if we could change B while leaving everything else fixed (as much as possible), the subject's experience would also change. If we can reliably alter experience in this way by manipulating B, then B is a *difference-maker* for P. Difference-makers for P are part of the explanation of P (subject to some further constraints below). Hence, we explain why the subject sees red rather than blue by noting the actual activity in B.

Interventions like that in 1b have already played an important role in our understanding of the brain. Simple contents of experience—features like colours, sounds, shapes or pains—can be affected in a variety of ways. Since Penfield's pioneering work (Penfield and Rasmussen 1950; Penfield and Jasper 1954), we have known that direct electrical stimulation of the cortex can produce a wide variety of experiences, with specific regions responsible for specific contents. Non-invasive stimulation (e.g. by TMS) can produce similar results, though with less specificity.

There is now considerable evidence that the contents of consciousness can be intervened upon in numerous systematic ways; for excellent reviews of the effect of neural stimulation on conscious states see Cohen and Newsome (2004) and Wu (2018). Notably, the effect of stimulation is not limited to simple phosphenes and the like, but can also include distortions of complex stimuli such as faces (Parvizi et al. 2012) and words (Hirshorn et al. 2016).

We motivated example 1a by discussing correlations between brain regions and experience. This suggests a link to the NCC project sketched above. Yet despite the name, we suggest that the search for NCC has actually been a search for difference-makers rather than for correlates as such. The notion of a 'content-specific NCC', e.g. is often glossed in explicitly difference-making terms (Koch et al. 2016; Boly et al. 2017). Koch et al. (2016, 208) claim that, were content-specific NCCs for face perception artificially activated, '...the participant should see a face even if none is present...'. Hence even if the primary evidence for NCCs is correlational (e.g. via neuroimaging) that evidence can be understood as revealing potential loci of intervention (Aru et al. 2012; Klein 2017).

In the same vein, recent debates (Boly et al. 2017; Odegaard et al. 2017) over whether the NCCs for visual experiences are more anterior or posterior fundamentally turn on which neural activity represents the working of the NCC versus the activity of mere precursors, background conditions and downstream effects (Aru et al. 2012; De Graaf et al. 2012). Each of these notions is well-studied in the interventionist literature. Mere background conditions, e.g. cannot be altered without altering many other facets of experience (and more besides).

Though familiar, interventions on contents raise a number of important points about explanation. Interventionist explanation is always contrastive: we explain why the world is one way *rather than another* by showing that some property is one way *rather than another*. As some authors have noted, this is also the case with explanations of conscious contents. Hohwy and Frith's (2004) proposed explanation for the feeling of conscious control, e.g. is explicitly meant to give an explanation of why a particular conscious feeling 'feels like the feeling of being in control, rather than like other closely related feelings'.

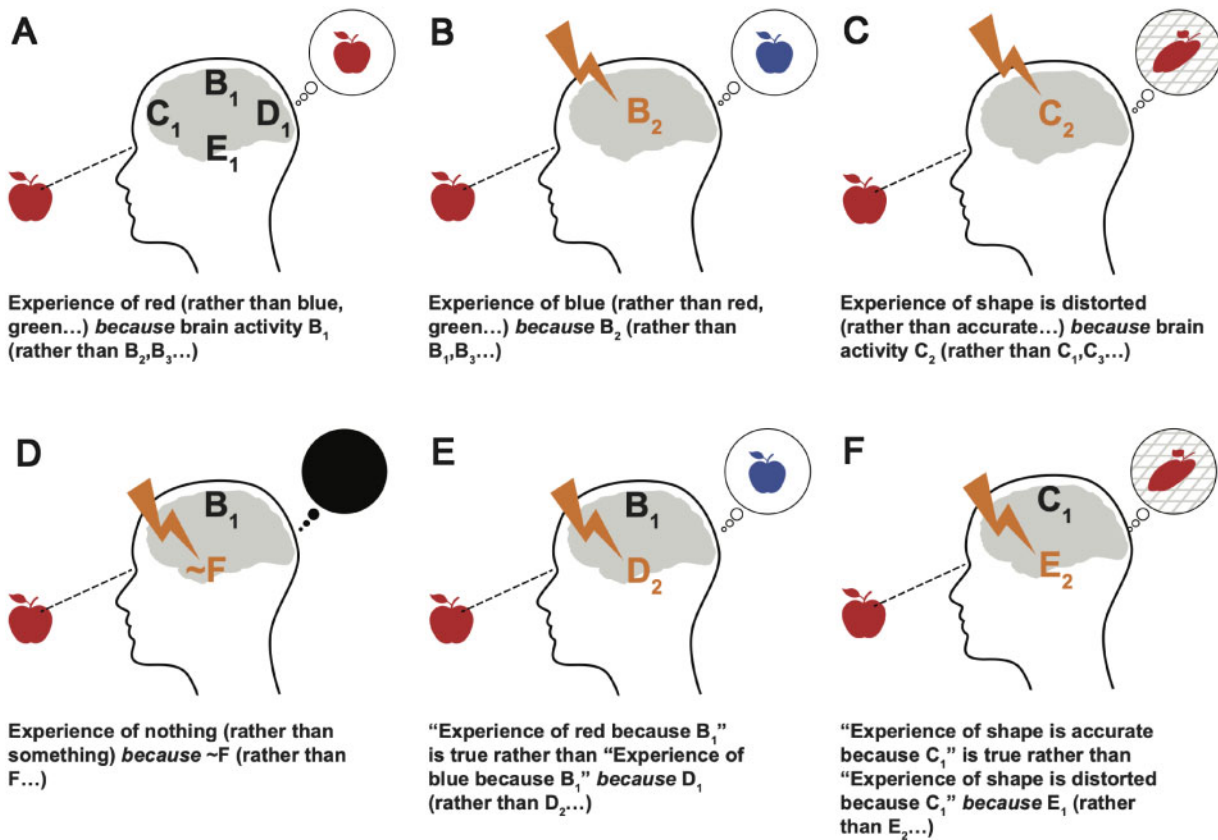


Figure 1. Different possible interventions of conscious processes. (a) The simple case, explaining a token experience. B_1 is a pattern of brain activation associated with seeing red, while C_1 , D_1 , E_1 , etc. are background conditions. The caption expresses a first-order phenomenal invariant. (b) An appropriate intervention on B will change the experienced colour given the same stimulus. This provides evidence for the claim in (a). (c) An intervention on a structural capacity. The metric of perceived space itself is distorted by an intervention of C , causing a variety of linked changes in experience of the stimulus. (d) A non-specific intervention on consciousness by eliminating a necessary condition. (e) A second-order intervention on a first-order invariant. D is part of what makes the laws in (a) and (b) true. Note that given the intervention on D , the same brain pattern B_1 that gives rise to a red experience in (a) gives rise to blue experience. (f) A similar intervention on structural capacities. Cases (a) and (b) are entirely compatible with property dualism, whereas the remaining cases would be problematic.

Thus the same phenomenon may receive different explanations when we consider different contrast classes (van Fraassen 1980; Hitchcock 1996; Woodward 2003). One important case where this happens is when a property depends on combinatorial properties of its realizer. What makes a pixel white rather than yellow is not the same as what makes it white rather than cyan. Because the property of whiteness depends on the combination of different underlying properties, different contrasts can require different explanations.

Interventions on conscious contents also give non-exclusive explanations. Non-exclusivity is a familiar feature in other special science explanations. Having genes for brown eyes explains why you have brown eyes (rather than some other colour). This is so even though those genes, on their own, cannot actually make an eye, and even though other genes make a difference to eye colour in a variety of ways (Sterelny and Kitcher 1988). Similarly, noting that activity in B is part of the explanation for why I experienced an apple as red (rather than blue) is not to exclude other brain regions from also having an effect on my phenomenal experience of colour. Unlike traditional formulations that talk about the NCC for an experience, therefore, interventionism does not assume that there needs to be a single NCC that is sufficient or necessary for a particular aspect of phenomenal awareness.

Interventions are not limited to simple contents. One may intervene on broader structural features of experience as well. That is, a single intervention might create correlated changes in a wide variety of contents and the relationships between them.

The literature contains examples of complex and subtle interventions on the global structure of experience, as illustrated in Fig. 1c. In a coda to *Awakenings*, Sacks (1999) describes the subjective distortions of space and time that are caused by the degeneration of the substantia nigra, suggesting an interesting relationship between the basal ganglia, the general perception of space and time and the specific motor impairments of Parkinsonian patients. Drugs such as dextromethorphan can produce striking alterations in the global perception of motion and time (Wolfe and Caravati 1995).

There are, of course, ways to alter structural features of experience even more profoundly. Anaesthetics eliminate consciousness altogether, and have been the subject of important work on the foundations of subjective experience (Alkire et al. 2008; Mashour and Alkire 2013; Klein and Barron 2016). As several authors have recently urged, the capacities that underlie consciousness, and hence the broader modes of variation possible, are probably numerous and heterogeneous (Bayne et al. 2016). That is, there is no simple, well-ordered scale of

‘degree’ of awareness. Instead, there are numerous dimensions along which conscious experience as a whole can vary.

Some interventions are crude. One might eliminate consciousness altogether by eliminating a necessary background condition (Fig. 1d). So, e.g. intervention on the claustrum appears to act as a kind of on-off switch for experience (Koubeissi et al. 2014). There are also many boring ways to cause mere unconsciousness. We are more interested in interventions that give a selective (Woodward Hopf and Bonci 2010; Griffiths et al. 2015) handle on phenomena we care about. Selective interventions are ones where there are many states of the control variable and many of the target, linked in a roughly one-to-one fashion, allowing for fine-grained control of the target. Similarly, one might also look for a systematic (Klein 2017) interventions, which allow fine-grained control over the *degree* to which a quantity changes.

A key feature of explanatory interventions is that they *reliably* alter their targets. (Formally speaking, the explanation relationship holds between event types, not tokens.) One way to put the point is that the relationship between *B* and *E* must be captured by an invariant generalization that holds across at least some changes in background conditions and some interventions on the target variables (Woodward 2003). We will call the special case of generalizations that connect brain processes and phenomenal processes ‘phenomenal invariants’.

Phenomenal invariants are like traditional psychophysical bridge laws (Davidson 1970), in that they connect brain to consciousness. Phenomenal invariants could be like fundamental laws of nature—i.e. they might hold over all background conditions and any interventions. This is arguably the picture embodied in versions of property dualism.

However, the logic of invariance permits more flexibility. Most special science generalizations are not invariant across all conditions or across all interventions on the properties they relate. Indeed, the move to interventionism within philosophy of science was triggered in part by examples showing that sufficiency and explanatory purchase come apart (Salmon 1989). Nevertheless, the fact that generalizations remain invariant under some conditions means that special science generalizations remain explanatorily useful. Indeed, experimental practice is often dedicated to finding both where an invariant holds and where it breaks down.

So too, we suggest, with phenomenal invariants—they might hold only over a limited range of conditions and interventions. Figure 1d shows a case where the same brain region fails to give rise to a red experience, because certain other conditions are not met. A difference-maker need not be sufficient for its effect, either individually or across all circumstances.

Higher-Order Explanation

We have said little that should be controversial. As we have suggested throughout, all parties to the contemporary debate agree that there are phenomenal invariants. The key question is whether they can be explained, and why (or why not). So, e.g. most property dualists think that the link between brain and consciousness is regular; the claim is merely that these phenomenal invariants are *sui generis* and irreducible (Chalmers 1996). By contrast, the physicalist ought to think that there is an explanation of phenomenal invariants.

The question is what would fit the bill. That is really a question about how generalizations themselves ought to be explained. The explanation of generalizations has received relatively less attention by fans of interventionism. What has been

done has mostly focused on mathematical and fundamental physical laws (Woodward 2003; Lange 2009; Gijsbers 2011, 2017), though see Rosenberg (2012) for an extension to economics.

In the case of consciousness, however, the investment in interventionism begins to pay dividends. The same principles that apply to the explanation of events can be extended to explain invariants themselves.

So far, we have focused on first-order phenomenal invariants. These connect brain processes to phenomenal processes. Schematically, a first-order phenomenal invariant like the relationship in Fig. 1b has the form:

$$G : B_1\{B_2, B_3 \dots B_n\} \rightarrow P_1\{P_2, P_3, \dots P_m\}.$$

Where the bits in brackets indicate possible contrast classes, and the arrow indicates that it is a proper change-relating generalization.

Now, *G* is a phenomenal invariant. If it is like most special science generalizations, it is contingent—i.e. some other invariant might hold in its place, e.g. something like:

$$G' : B_1\{B_2, B_3 \dots B_n\} \rightarrow P_2\{P_3, P_4, \dots P_1\}.$$

In *G'*, the very same brain process gives rise to a different phenomenal experience. We take the relevant sense of possibility here to be nomological: i.e. it might be possible, in our world, to change the phenomenal invariant cited.

What does it take to explain why *G* holds *rather than G'*? By parallelism, the answer is a second-order phenomenal invariant of the form:

$$\mathbb{G} : B_1^*\{B_2^*, B_3^*, \dots B_n^*\} \rightarrow G\{G', G'', \dots\}.$$

Like first-order phenomenal invariants, \mathbb{G} is change-relating: it shows how interventions on a brain region would affect the world. Rather than an effect on a phenomenal process, however, the effect is on a phenomenal invariant—changes to *B** change the whole *B*-to-*P* relationship.

Figure 1e and f illustrates the sorts of interventions that might be relevant. One can intervene to change both phenomenal invariants involving simple contents and more complex invariants that cover structural relationships. Of course, even higher-order phenomenal invariants are possible too—one could have a third-order invariant that alters how a second-order invariant works, and so on. For now, second-order invariants will be plenty.

An important feature of the account is that we can get evidence for second-order invariants directly—i.e. without further underpinnings by theory. Interventions like those depicted in Figure 1e and f show how second-order invariants can be discovered from the bottom-up.

First-order invariant generalizations concern difference-makers with a direct link to phenomenal effects. They are the analogue of the traditional psychophysical bridge laws, though with the important differences noted above. Second-order invariant generalizations, on the other hand, can cover a wide variety of processes, including traditionally non-phenomenal ones. A second-order invariant generalization shows how to change the relationship between a brain process and its phenomenal outcome, and the interventions that make that change possible need not necessarily involve ones with direct phenomenal consequences.

Hence one place to look for second-order invariants is by looking for interventions that change experiences by affecting

the cognitive capacities that form the background and context for phenomenal experience. The capacities necessary for conscious experience might be things like the functions supporting conscious experience such as selective attention, integrative and interactive processing of exteroceptive and interoceptive information, a unified spatial and temporal framework for sensory information, or unlimited associative memory (Ginsburg and Jablonka 2007; Merker 2007; Ginsburg and Jablonka 2010; Barron and Klein 2016). These are not themselves phenomenal, but they can be involved in preserving and maintaining first-order phenomenal invariants.

Individual conscious contents are also shaped and given meaning by the structural features of experience that order and contain them. For example, we experience visual and auditory sensations as occurring within a common, external space. Yet while these structural features are the conditions for the possibility of conscious experience, they are arguably not objects of experience themselves (Kant 1999). Hence a particular content like the experience of an apple is bound up in structural features of phenomenal experience, which in turn seem to depend on broader neural capacities for information-processing in both phenomenal and non-phenomenal domains.

This means that the very same brain activation might give rise to a different sensation or systematic alterations in broader underlying context. In the phenomenon known as ‘*pain asymbolia*’, patients with anterior insula damage will report that they continue to feel pain but no longer care about it (Schilder and Stengel 1931). Similar effects occur with a range of other dissociative drugs (Keats and Beecher 1950). There is debate about whether this effect is due to a sensory-limbic dissociation (Grahek 2007) or to a general breakdown in processes of bodily ownership and concern (Klein 2015). Either way, there is alteration to the character of individual sensations by changes in the background conditions of experience.

Of course, it is a working hypothesis that there are second-order invariants, and that there are systematic, theory-based ways to generate and discover them. That could be wrong, in which case an experimentalist project would be practically useful but theoretically fruitless. We take it, in other words, that this is a working hypothesis that can be falsified, which is a virtue of an empirical programme.

On the other hand, it might be that interventions on cognitive capacities and the like reveal systematic, specific ways to intervene on first-order phenomenal invariants. But of course, on the interventionist picture, what it is to explain a phenomenon is to demonstrate how it can be made to vary in replicable, systematic ways. So showing how first-order phenomenal invariants can be made to alter in ways that fall under non-phenomenal, non-*sui generis* second-order order invariants just is what it takes to explain the first-order invariants. That is what a materialist solution to the hard problem demanded.

Is That All There Is?

You might think that this is not quite what you were looking for. What the interventionist programme offers is explanations of a wide variety of first-order phenomenal invariants. If successful, we will be able to explain why this activity in this brain region causes the experience of a red apple (rather than blue), why activity in that region makes you experience a face *this way* rather than *that way*, and so on.

Yet—the objection goes—this leaves something out. The task of the hard problem was not to explain many first-order invariants, though those explanations will be interesting. It is to

do something more philosophically heavy: to explain why there is phenomenal experience at all! As we stressed in the setup, the hard problem was about explaining consciousness—not why the generalizations connecting brain to experience have one feature rather than another. Furthermore, contrastive explanation is by its nature pluralist: there will be not one grand explanation but many interlinking explanations. That might feel like something of a letdown.

The objection admits of two readings, one scientific and one more subjective. We think that, properly spelled out, neither sticks.

First, one might construe the objection scientifically, as a residual question that the interventionist programme would not be able to answer. Contrastively, it would be something like: why is there phenomenal consciousness at all, *rather than not*? Note that this has to be read as a general question—there are a great number of first-order phenomenal invariants concerning why, on particular occasions, one is or is not aware of something.

So phrased, however, we suggest that the objection embodies something of a misconception about how inter-theoretic explanation works in complex domains. The picture—which is present in many older accounts of inter-theoretic reduction—is one on which there are two completed sciences, each of which posits a set of primitives, and the goal is to line up one science with the other.

It is questionable whether even canonical examples of inter-theoretic reduction fit this mould (Sklar 1974; Klein 2009). Regardless, the history of special science explanations is often far messier. Properties initially taken as scientifically primitive or simple often turn out to elaborate and fragment as they are investigated further. Inter-theoretic constraints often serve to accelerate this process, with the ontology in one science constraining that of the other, and vice versa, in an iterative fashion (Wimsatt 1974). This, in turn, fixes what used to seem like insoluble mysteries: what seemed like a deep question about a *sui generis* property dissolves into a bunch of tractable experimental questions.

The history of science provides numerous optimistic parallels. A closer look reveals that what initially appear to be grand, singular explanatory projects always end up dissolving into an array of specific, contrastive explanations as science advances. So, e.g. much of early chemistry was preoccupied with how to explain the chemical role of ‘*water*’, given the specific assumption that water was an element (Chang 2012). Ontologies that included elemental water could get quite far, and, as Chang points out, could have advanced even further. What broke the impasse was not independent advances in atomic chemistry that were subsequently used to explain the properties of water. Rather, these puzzles were untangled by the development of specific manipulations of the properties of water—by, e.g. investigating the properties of solutes, or advances in electrolysis. These manipulations in turn helped inform higher-order generalizations that were ultimately key to the development of atomic chemistry, and the upending of water as a basic element.

Similarly, in the 18th century, there was a grand philosophical challenge to explain ‘*Life*’ (Nassar 2016). Considered as such, little progress could be made. Vitalism remained plausible. The advance of physiology in the 18th century did not attempt to explain life as a whole. Rather, it explained why this inorganic process could give rise to urea, why that process kept blood pH within reasonable limits, while that process cleared carbon dioxide rather than letting it accumulate, and so on (Bernard

1865/1949). The march of progress ends up dissolving the original grand problem into an array of contrastive explanations, leaving even the project of defining ‘Life’ as a questionable one (Machery 2012). We have not explained ‘Life’ as it preoccupied the early modern philosophers. Instead, we can explain a great variety of things about living beings.

Similarly so, many have argued, with *consciousness* (Dennett 1996; Seth 2016). Couched in the terminology of philosophy of explanation, one might say that this formulation of the hard problem should be rejected rather than answered (van Fraassen 1980). One rejects an explanatory demand when it contains a false presupposition. If you ask why red mercury can initiate nuclear fusion, all I can tell you is that there is no such thing. That does not answer your question: it says that there is a problem with the question itself.

Indeed, chemistry and life offer a useful model for an experimentalist rejection of questions that once seemed sensible. The problem comes from considering a property as primitive; the solution comes from realizing that the problematic property is a feature of a diverse set of underlying mechanisms that can combine in a variety of different ways, including ways that seemed primitive. Hence to ask ‘Why is there consciousness at all, rather than not?’ is to assume that there is a single thing and a single answer. That expectation is at the heart of the hard problem. There are many useful questions that one *can* answer, of course, and the answering of which is part of the rejection of the simple question.

The second way of reading the objection focuses on whether higher-order generalizations are satisfying explanations of consciousness. Even granting the parallel to other scientific endeavours, there is something about the hard problem that feels different than other problems. Chalmers (2018) has recently dubbed this the ‘*meta-problem of consciousness*’. A satisfying solution to the hard problem ought to explain why it seemed like there was a hard problem in the first place—why first-order invariants seem arbitrary and inexplicable, even if they are not. Many otherwise promising accounts clearly fail to fit the bill.

We think this is a serious challenge. We are fond of a standard sort of answer. There is a view, tracing back at least to Leibniz, on which the apparent simplicity and arbitrariness of conscious processes is merely an introspective confusion about a complex underlying process (Hilbert 1987; Armstrong 1997; Pettit 2003). As Lashley famously put it: ‘No activity of mind is ever conscious... There is order and arrangement, but there is no experience of the creation of that order’ (Lashley et al. 1960). The hard problem arises because we lack access to the relevant goings-on. There may be other sources of trouble as well, such as our relatively limited capacity for introspection and discussion of our conscious processes compared to the richness of conscious experience itself (Block 2011).

Each of these mechanisms is a fact about us and how we are constituted, rather than a deep metaphysical feature of the world. So our subjective experience is underpinned by a great number of mechanisms to which we have no conscious access, and that are not themselves represented in conscious experience. As we are aware only of the products of a complex mechanism and not its actual workings, we feel an arbitrariness of, and passivity towards, those products. The unconscious workings that give rise to conscious experience do not require effort of will and do not admit of first-person control. That is why the details of conscious processes feel arbitrary: subjectively, they simply appear out of nowhere.

Yet simply knowing all of this does not, by itself, make conscious experience feel any less arbitrary. That is the sense in which the hard problem is a unique scientific problem: being told the answer will not remove the sense of mystery.

That tangle is a problem with us, stemming from how we are constituted. So phrased, we think there is a straightforward solution. We think that the meta-problem demands interventions that also change our relationship to our own phenomenal properties—i.e. via *self-intervention*.

Interventions on brain processes have both an objective and a subjective component. By intervening on brains, we do not simply discover that certain experiences can be evoked, or that they depend on certain interventions. The first-person, subjective experience of that intervention is critical as well.

This is not just proof of principle, though the proof of principle is important. (It is one thing to read about the experience of alien hand syndrome, and quite another to feel your fingers jump around under the influence of TMS.) We believe that by feeling how first-order invariants can be altered by altering brain activity the impression of arbitrariness should vanish. The more systematic and specific we can self-intervene, the less arbitrary things will feel.

One does not need to be able to self-intervene on every first-order phenomenal invariant. Good thing too, for that would be impossible—many interventions would undermine the ability to introspect. The role of self-intervention is primarily to remove the air of arbitrariness that falsely attends explanations of first-order phenomenal invariants. A little is likely to go a long way.

Thus, self-manipulation of brain activity has the unique possibility not just to solve the hard problem, but to fix the passivity that leads to both the hard and the meta-problem. The point of interventions is to give us points of mastery over the world (Campbell 2007, 2010). Self-mastery will be, and probably must be, the key to pushing past the lack of understanding that holds back effective research.

Making Progress

We recognize that our proposal is an unusual solution. Self-manipulation in particular has a slightly dodgy history. What we can do now should be seen, at best, as necessary but possibly remote steps in a long process. Nevertheless, to make progress, we will ultimately need safe, specific, selective techniques that allow us to intervene on experience in awake adults. It is worth canvassing what is already available. At present the possibilities are either pharmacological, invasive or non-invasive. We have noted some possibilities above, and we consider each in turn.

Pharmacological intervention is the most familiar and accessible way to intervene upon consciousness. Anaesthetics remove consciousness altogether, and the specific ways and mechanisms by which consciousness breaks down already provides useful data about the capacities underlying experience (Alkire et al. 2008; Mashour and Alkire 2013).

There is an old tradition by which more specific interventions via psychedelic drugs have been thought to reveal interesting structural features about experience. There has been a recent revival of interest in psychedelics given their promising results in treating conditions like post-traumatic stress disorder (Oehen et al. 2013; Amoroso and Workman 2016; Mithoefer et al. 2018). That said, we think there is serious danger of repeating the mistakes of the past. We should be wary of returning to the kind of uncritical pharmacological investigations that were

popular among an earlier generation of researchers (Jay 2009; Lattin 2010). Some authors have been tempted to claim that the psychedelic experience itself is interesting precisely because it allows normally unconscious properties of the mind to be made manifest as objects of consciousness (Letheby 2015). This is an old idea, embodied in the etymology of ‘psychedelic’ itself. We are sceptical. Despite decades of citizen science, we note few lasting contributions of such work to modern understanding of cognitive mechanisms.

Part of the problem is that psychedelics tend to have widespread and complex effects on consciousness. Less common drugs with more limited effects may be more useful. We have suggested that second-order interventions that change the relationship between brain activity and the corresponding experience will be particularly useful. The literature contains tantalizing suggestions in this regard. For example, reports suggest that low doses of diisopropyltryptamine (DiPT) have effects primarily limited to non-linear distortions of audition (Shulgin 2000). Limited and well-defined phenomena may also be fruitfully investigated, as e.g. in work done using LSD to investigate the central mechanisms of binocular rivalry (Carter and Pettigrew 2003).

Invasive interventions involving direct electrical stimulation of the brain have been important to understand conscious function (Penfield and Rasmussen 1950). Invasive work presents obvious ethical and practical concerns, and so is done concurrent with some medical need. Much of the direct intervention work has focused on the effects of cortical stimulation on the contents of consciousness. However, there is increasing evidence that direct stimulation of the posterior cingulate/precuneus can produce more profound alterations in global experience (Herbet et al. 2014; Balestrini et al. 2015; Herbet et al. 2015). This would be consonant with these regions’ purported role in consciousness and mediating cortical-subcortical interactions (Vogt and Laureys 2005; Cavanna and Trimble 2006). Again, we note that these broader alterations might give clues to second-order invariants that give us a better handle on the first-order relationships between brain and experience.

As for subcortical interventions, deep brain stimulation (DBS) has shown intriguing evidence of effects on consciousness. Much of this evidence takes the form of alleviation (Krack et al. 2010; Lyons 2011) or induction (Bejjani et al. 1999) of psychiatric conditions such as obsessive-compulsive disorder and depression. Thalamic DBS has also led to promising improvements in minimally conscious patients (Schiff et al. 2007). The variety of possible stimulation parameters, and the variability of results between microstimulation and direct electrical stimulation (Vincent et al. 2016) suggests a fruitful experimental programme in this area. We note that many case reports present no or only minimal data about a patients’ subjective experience, even when this would clearly be accessible. We suggest that these data ought to be more routinely and systematically collected.

Finally, non-invasive brain stimulation such as transcranial electrical stimulation (tES) may avoid the practical problems associated with invasive interventions. There have been initial indications that tES can improve responsiveness of patients in Minimally Conscious States (Thibaut et al. 2014). Perhaps the most interesting applications of tES, using either DC or AC current, is the possibility of entraining underlying circuits and thereby altering temporal dynamics of brain activation (Filmer et al. 2014; Tavakoli and Yun 2017). tES has had problems showing specificity and replicability, but recent techniques using

EEG/MEG to guide stimulation timing (Thut et al. 2017) may help ameliorate these concerns.

Most of the existing interventions we have discussed are still relatively broad and uncontrolled. The ability to make more systematic interventions would make subjective experiences seem less like passive and fleeting epiphenomena; they could be controlled, evoked and altered at will. Ultimately, the requirement for specific interventions will demand developing new ways to intervene on the brain.

Invasive interventions also occur in research on brain-machine interfaces, though it is early days for this field. The current focus is on developing devices that can interact with neural circuits in such a way that they can become part of the system of information representation (Clark 1995); the aim being to supplement or replace memory, or even add new information representations (Berger et al. 2011; Deadwyler et al. 2013; Deadwyler et al. 2017).

Much of the work on developing new forms of brain-machine interface is currently happening with animals. This is the norm for experimental interventionist neuroscience. It is unethical to develop new methods on humans, but the reality of the deep homology of brain system functions across vertebrates (Striedter 2005), and of neuron functions across most animal phyla (Kristan 2016), means that methods developed in one species can usually be translated (with informed modifications) to another.

There is, however, a unique tension in using animal systems to study the nature of conscious experience. There remains a lively debate around which animals have any conscious experience at all, precisely because we do not know how neural circuits support conscious experience. Furthermore, solving the meta-problem ultimately requires self-intervention, so animal models can only ever do part of the job. That said, we envision research with animal models playing a key role for developing the interventionist tools, methods and approaches needed for an experimental investigation of the hard problem in humans. Indeed, even very simple animals such as insects might provide a useful test-bed for developing more complex interventions (Barron and Klein 2016; Klein and Barron 2016).

Finally, we acknowledge that self-stimulation is likely to raise unique ethical issues. The present proposal can be seen in the spirit of self-experimentation that characterized experimental medicine in the first half of the 20th century (Altman 1998) and was a vital part of the development of novel psychedelics (Shulgin 2000). In a popular essay, Haldane (1927) suggested that self-experimentation was useful precisely because it had fewer ethical considerations than animal experiments. Some have defended self-experimentation precisely because of the personal insight it brings and for the personal commitment it requires (Dresser 2014).

Yet the danger of self-experimentation can be considerable. Furthermore, in addition to the immediate danger to the experimenter, there are broader knock-on effects that could be ethically problematic. The Stanford neuroscientist Bill Newsome, e.g. claims that self-stimulation would be scientifically valuable, but could set up a problematic slippery slope by encouraging early-career researchers to make risky choices (Singer 2006). Clearly, ethical guidelines for self-experimentation must take into account this additional moral hazard.

Conclusion: Fixing the Hard Problem

We have outlined an ambitious programme for solving the hard problem. The hard problem of consciousness has two roots: an

outdated philosophy of science, and a deep (but not insuperable) limitation in our own ability to understand the roots of our experiences. Having identified these, neuroscientists must fix those shortcomings. This will require direct intervention and a mix of third-person and first-person techniques.

A similar process allowed us to make progress on other seemingly insoluble scientific problems. Similarly, we envision, with consciousness. Successful interventionist research projects will alter and vary the relationship between brain activity and subjective experience. This will elucidate important mechanisms, and allow ever-finer control of experience. In the limit case, we will find consciousness just as grand as, but no more mysterious than, water or life.

Finally, we re-emphasize that this is a research programme that is fundamentally falsifiable. We might find that there are no systematic ways to intervene on first-order phenomenal invariants: they are like the brute laws of fundamental physics. Were the evidence to go that way, then non-physicalist theories of consciousness would gain plausibility.

Such a project obviously faces a host of practical problems. We do not pretend it will be easy. Many of the techniques and frameworks for brain intervention that will be required are only dimly understood at present. But unless we embark on the kind of interventionist neuroscience programme we have described the appearance of a hard problem will persist. Our discoveries about consciousness will always have a whiff of the arbitrary. The open question—why this?—will linger in the air.

Yet we think it is worth being optimistic. The idea that the hard problem might be a practical problem rather than a philosophical one has an unexpected pedigree. When Nagel (1974, 447) argued that we do not know what it is like to be a bat, his point was not to argue against physicalism. Though often overlooked, Nagel closes his discussion with a positive proposal. Part of our difficulty in understanding consciousness, he says, is reliance on imagination when we try to take up the point of view of another subject. Imagination is an inherently limited faculty. Hence, Nagel (1974, 449) tells us, his argument should be seen as ‘a challenge to form new concepts and devise a new method’ of approaching experience. We agree. We advocate for tackling consciousness directly, by intervening at its roots.

Conflict of interest statement. None declared.

Acknowledgements

This paper was funded by Australian Research Council Grant FT140100422 (to C.K.) and FT140100452 (to A.B.B.). Thanks to François Kammerer, Daniel Stoljar, several anonymous reviewers and audiences at ANU, University of Melbourne and the Blackheath Philosophy Forum for helpful feedback on earlier drafts.

References

- Alkire MT, Hudetz AG, Tononi G. Consciousness and anesthesia. *Science* 2008;**322**:876–80.
- Altman LK. *Who Goes First?: The Story of Self-experimentation in Medicine*. Los Angeles: University of California Press, 1998.
- Amoroso T, Workman M. Treating posttraumatic stress disorder with MDMA-assisted psychotherapy: a preliminary meta-analysis and comparison to prolonged exposure therapy. *J Psychopharmacol* 2016;**30**:595–600.
- Armstrong D. What is consciousness? In: Block N, Owen Flanagan O, Güzeldere G (eds.) *The Nature of Consciousness: Philosophical Debates*. Cambridge: MIT Press, 1997, 721–8.
- Aru J, Bachmann T, Singer W et al. Distilling the neural correlates of consciousness. *Neurosci Biobehav Rev* 2012;**36**:737–46.
- Balestrini S, Francione S, Mai R et al. Multimodal responses induced by cortical stimulation of the parietal lobe: a stereo-electroencephalography study. *Brain* 2015;**138**:2596–607.
- Barron AB, Klein C. What insects can tell us about the origins of consciousness. *Proc Natl Acad Sci USA* 2016;**113**:4900–8.
- Bayne T, Hohwy J, Owen AM. Are there levels of consciousness? *Trends Cogn Sci* 2016;**20**:405–13.
- Bejjani B-P, Damier P, Arnulf I et al. Transient acute depression induced by high-frequency deep-brain stimulation. *N Engl J Med* 1999;**340**:1476–80.
- Berger TW, Hampson RE, Song D et al. A cortical neural prosthesis for restoring and enhancing memory. *J Neural Eng* 2011;**8**: 046017.
- Bernard C. *An Introduction to the Study of Experimental Medicine: Henry Schuman*. New York: Henry Schuman, Inc, 1865/1949.
- Block N. Perceptual consciousness overflows cognitive access. *Trends Cogn Sci* 2011;**15**:567–75.
- Boly M, Massimini M, Tsuchiya N et al. Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *J Neurosci* 2017;**37**: 9603–13.
- Campbell J. An interventionist approach to causation in psychology. In: Gopnik A, Schulz L (eds.) *Causal Learning: Psychology, Philosophy and Computation*. Oxford: Oxford University Press, 2007, 58–66.
- Campbell J. II-control variables and mental causation. *Proc. of the Aristotelian Society*, 2010, 15–30.
- Carter OL, Pettigrew JD. A common oscillator for perceptual rivalries? *Perception* 2003;**32**:295–305.
- Cavanna AE, Trimble MR. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 2006;**129**: 564–83.
- Chalmers D. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, 1996.
- Chalmers DJ. On the search for the neural correlate of consciousness. *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates* 1998;**2**:219.
- Chalmers DJ. Consciousness and its place in nature. In: Stich SP, Warfield TA (eds.), *Blackwell Guide to the Philosophy of Mind*. Hoboken: Blackwell, 2003, 102–42.
- Chalmers DJ. The meta-problem of consciousness. *J Consciousness Stud* 2018;**25**:6–61.
- Chang H. *Is Water H2O?: Evidence, Realism and Pluralism*. New York: Springer Science & Business Media, 2012.
- Clark A. Moving minds: situating content in the service of real-time success. *Philos Perspect* 1995;**9**:89–104.
- Cohen MR, Newsome WT. What electrical microstimulation has revealed about the neural basis of cognition. *Curr Opin Neurobiol* 2004;**14**:169–77.
- Craver CF. *Explaining the Brain*. New York: Oxford University Press, 2007.
- Crick F, Koch C. Towards a neurobiological theory of consciousness. *Semin Neurosci* 1990;**2**:263–75.
- Davidson D. Mental events. In: Foster L, Swanson JW (eds.), *Essays on Actions and Events*. Oxford: Clarendon Press, 1970, 207–24.
- De Graaf TA, Hsieh P-J, Sack AT. The ‘correlates’ in neural correlates of consciousness. *Neurosci Biobehav Rev* 2012;**36**: 191–7.

- Deadwyler SA, Berger TW, Sweatt AJ et al. Donor/recipient enhancement of memory in rat hippocampus. *Front Syst Neurosci* 2013;1:120.
- Deadwyler SA, Hampson RE, Song D et al. A cognitive prosthesis for memory facilitation by closed-loop functional ensemble stimulation of hippocampal neurons in primate brain. *Exp Neurol* 2017;287:452–60.
- Dennett D. Facing backwards on the problem of consciousness. *J Consciousness Stud* 1996;3:4–6.
- Dennett DC. *Consciousness Explained*. Boston: Little, Brown, & Co., 1991.
- Dresser R. Personal knowledge and study participation. *J Med Ethics* 2014;40:471–4.
- Feinberg TE, Mallatt JM. *The Ancient Origins of Consciousness: How the Brain Created Experience*. Cambridge, MA: MIT Press, 2016.
- Filmer HL, Dux PE, Mattingley JB. Applications of transcranial direct current stimulation for understanding brain function. *Trends Neurosci* 2014;37:742–53.
- Gijssbers V. *Explanation and Determination*. PhD thesis. Leiden University, 2011.
- Gijssbers V. A quasi-interventionist theory of mathematical explanation. *Logique Anal* 2017;237:47–66.
- Ginsburg S, Jablonka E. The transition to experiencing: the evolution of associative learning based on feelings. *Biol Theor* 2007;2: 231–43.
- Ginsburg S, Jablonka E. The evolution of associative learning: a factor in the Cambrian explosion. *J Theor Biol* 2010;266:11–20.
- Ginsburg S, Jablonka E. *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. Cambridge: MIT Press, 2019.
- Grahek N. *Feeling Pain and Being in Pain*. Cambridge: MIT Press, 2007.
- Griffiths PE, Pocheville A, Calcott B et al. Measuring causal specificity. *Philos Sci* 2015;82:529–55.
- Haldane JBS. On being one's own rabbit. In: *Possible Worlds*. London: Phoenix Library, 1927, 107–19.
- Herbet G, Lafargue G, de Champfleury NM et al. Disrupting posterior cingulate connectivity disconnects consciousness from the external environment. *Neuropsychologia* 2014;56:239–44.
- Herbet G, Lafargue G, Duffau H. The dorsal cingulate cortex as a critical gateway in the network supporting conscious awareness. *Brain* 2015;139:e23.
- Hilbert DR. *Color and Color Perception: A Study in Anthropocentric Realism*. Stanford: Center for the Study of Language and Information, 1987.
- Hirshorn EA, Li Y, Ward MJ et al. Decoding and disrupting left midfusiform gyrus activity during word reading. *Proc Natl Acad Sci USA* 2016;113:8162–7.
- Hitchcock CR. The role of contrast in causal and explanatory claims. *Synthese* 1996;107:395–419.
- Hohwy J, Bayne T. The neural correlates of consciousness: causes, confounds and constituents. In: Miller SM (ed.), *The Constitution of Phenomenal Consciousness - Toward a Science and Theory*. Netherlands: John Benjamins Publishing Company, 2015, 155–76.
- Hohwy J, Frith C. Can neuroscience explain consciousness? *J Consciousness Stud* 2004;11:180–98.
- Irvine E. *Consciousness as a Scientific Concept: A Philosophy of Science Perspective*. New York: Springer, 2013.
- Jay M. *The Atmosphere of Heaven: The 1799 Nitrous Oxide Researches Reconsidered*. Notes and Records of the Royal Society 2009;63:297–309.
- Kammerer F. *The Illusion of Conscious Experience*. Synthese: online first. DOI 10.1007/s11229-018-02071-y. 2019.
- Kant I. *Critique of Pure Reason: The Cambridge Edition of the Works of Immanuel Kant*. Cambridge: Cambridge University Press, 1999.
- Keats AS, Beecher HK. Pain relief with hypnotic doses of barbiturates and a hypothesis. *J Pharmacol Exp Ther* 1950;100:1–13.
- Klein C. Reduction without reductionism: a defence of Nagel on connectability. *Philos Q* 2009;59:39–53.
- Klein C. What pain asymbolia really shows. *Mind* 2015;124: 493–516.
- Klein C. Brain regions as difference-makers. *Philos Psychol* 2017; 30:1–20.
- Klein C, Barron AB. Insects have the capacity for subjective experience. *Animal Sentience* 2016:100.
- Koch C, Massimini M, Boly M et al. Neural correlates of consciousness: progress and problems. *Nat Rev Neurosci* 2016;17: 307–21.
- Koubeissi MZ, Bartolomei F, Beltagy A et al. Electrical stimulation of a small brain area reversibly disrupts consciousness. *Epilepsy Behav* 2014;37:32–5.
- Krack P, Hariz MI, Baunez C et al. Deep brain stimulation: from neurology to psychiatry? *Trends Neurosci* 2010;33:474–84.
- Kristan WB. Early evolution of neurons. *Curr Biol* 2016;26: R937–80.
- Lange M. *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*. Oxford: Oxford University Press, 2009.
- Lashley KS, Beach FA, Hebb DO et al. *The Neuropsychology of Lashley: Selected Papers of KS Lashley*. New York: McGraw Hill, 1960.
- Lattin D. *The Harvard Psychedelic Club: How Timothy Leary, Ram Dass, Huston Smith, and Andrew Weil Killed the Fifties and Ushered in a New Age for America*. New York: Harper Collins, 2010.
- Letheby C. The philosophy of psychedelic transformation. *J Consciousness Stud* 2015;22:170–93.
- Levine J. Materialism and qualia: the explanatory gap. *Pac Philos Q* 1983;64:354–61.
- Lyons MK. Deep brain stimulation: current and future clinical applications. In: *Mayo Clinic Proceedings*, 2011, 662–72.
- Machery E. Why I stopped worrying about the definition of life... and why you should as well. *Synthese* 2012;185:145–64.
- Mashour GA, Alkire MT. Consciousness, anesthesia, and the thalamocortical system. *Anesthesiology* 2013;118:13–5.
- Merker B. Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav Brain Sci* 2007;30:63–81.
- Mithoefer MC, Mithoefer AT, Feduccia AA et al. 3, 4-methylenedioxymethamphetamine (MDMA)-assisted psychotherapy for post-traumatic stress disorder in military veterans, firefighters, and police officers: a randomised, double-blind, dose-response, phase 2 clinical trial. *Lancet Psychiatry* 2018;5:486–97.
- Nagel T. What is it like to be a bat? *Philos Rev* 1974;83:435–50.
- Nassar D. Analogical reflection as a source for the science of life: Kant and the possibility of the biological sciences. *Stud Hist Philos Sci* 2016;58:57–66.
- Odegaard B, Knight RT, Lau H. Should a few null findings falsify prefrontal theories of conscious perception? *J Neurosci* 2017;37: 9593–602.
- Oehen P, Traber R, Widmer V et al. A randomized, controlled pilot study of MDMA (\pm 3,4-methylenedioxymethamphetamine)-assisted psychotherapy for treatment of resistant, chronic Post-Traumatic Stress Disorder (PTSD). *J Psychopharmacol* 2013;27: 40–52.
- Parvizi J, Jacques C, Foster BL et al. Electrical stimulation of human fusiform face-selective regions distorts face perception. *J Neurosci* 2012;32:14915–20.
- Pearl J. *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press, 2000.

- Penfield W, Jasper H. *Epilepsy and the Functional Anatomy of the Brain*. London: J.&A. Churchill, Ltd, 1954.
- Penfield W, Rasmussen T. *The Cerebral Cortex of Man: A Clinical Study of Localization of Function*. New York: Macmillan, 1950.
- Pettit P. Looks as powers. *Philos Issues* 2003;13:221–52.
- Rosenberg A. Why do spatiotemporally restricted regularities explain in the social sciences? *Br J Philos Sci* 2012;63:1–26.
- Sacks O. Parkinsonian space and time. In: *Awakenings*. New York: Vintage Books, 1999, 339–49.
- Salmon W. *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press, 1989.
- Schiff ND, Giacino JT, Kalmar K et al. Behavioural improvements with thalamic stimulation after severe traumatic brain injury. *Nature* 2007;448:600.
- Schilder P, Stengel E. Asymbolia for pain. *Arch Neurol Psychiatry* 1931;25:598–600.
- Seth AK. *The Real Problem*. Aeon, 2016. <https://aeon.co/essays/the-hard-problem-of-consciousness-is-a-distraction-from-the-realone> (5 February 2018, date last accessed).
- Shulgin A. *Tryptamines I Have Known and Loved*. Berkeley: Mind Books, 2000.
- Singer E. *Big Brain Thinking*. MIT Technology Review, 2006. <https://www.technologyreview.com/s/405296/big-brain-thinking/> (10 April 2020, date last accessed).
- Sklar L. Thermodynamics, statistical mechanics and the complexity of reductions. *PSA: Proc Biennial Meeting Philos Sci Assoc* 1974;1974:15–32.
- Sterelny K, Kitcher P. The return of the gene. *J Philos* 1988;85:339–61.
- Striedter GF. *Principles of Brain Evolution*. Sunderland, MA: Sinauer Associates, 2005.
- Tavakoli AV, Yun K. Transcranial alternating current stimulation (tACS) mechanisms and protocols. *Front Cell Neurosci* 2017;11:214.
- Thibaut A, Bruno M-A, Ledoux D et al. tDCS in patients with disorders of consciousness Sham-controlled randomized double-blind study. *Neurology* 2014;82:1112–8.
- Thut G, Bergmann TO, Fröhlich F et al. Guiding transcranial brain stimulation by EEG/MEG to interact with ongoing brain activity and associated functions: a position paper. *Clin Neurophysiol* 2017;128:843–57.
- van Fraassen BC. *The Scientific Image*. New York: Oxford University Press, 1980.
- Vincent M, Rossel O, Hayashibe M et al. The difference between electrical microstimulation and direct electrical stimulation—towards new opportunities for innovative functional brain mapping? *Rev Neurosci* 2016;27:231–58.
- Vogt BA, Laureys S. Posterior cingulate, precuneal & retrosplenial cortices: cytology & components of the neural network correlates of consciousness. *Prog Brain Res* 2005;150:205–17.
- Wimsatt WC. Reductive explanation: a functional account. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1974, 671–710.
- Wolfe TR, Caravati EM. Massive dextromethorphan ingestion and abuse. *Am J Emerg Med* 1995;13:174–6.
- Woodward J. Explanation and invariance in the special sciences. *Br J Philos Sci* 2000;51:197–254.
- Woodward J. *Making Things Happen*. New York: Oxford University Press, 2003.
- Woodward J. Causation in biology: stability, specificity, and the choice of levels of explanation. *Biol Philos* 2010;25:287–318.
- Woodward Hopf F, Bonci A. Dnmt3a: addiction’s molecular forget-me-not? *Nat Neurosci* 2010;13:1041–3.
- Wu W. The neuroscience of consciousness. In: *The Stanford Encyclopedia of Philosophy: Metaphysics Research Lab*. Stanford University, 2018.