How Experimental Neuroscientists Can Fix the Hard Problem of Consciousness

Colin Klein[1] and Andrew B. Barron[2]

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

Abstract

Contemporary neuroscience is making rapid progress in understanding how our nervous system perceives, processes, and responds to stimuli, but we currently have no clear understanding of why stimuli we perceive are conscious. Why should perceiving blue yield for a human the subjective mental experience of blueness? Philosophers have termed this the *hard problem* of consciousness. Many argue that experimental neuroscience can be of no use for tackling the hard problem. We disagree. Here we argue that the only way to understand how the brain gives rise to conscious experience is through an experimental program of intervention on the brain that directly and selectively varies different aspects of conscious experience. Consciousness seems arbitrary because the roots of our conscious experience are currently inaccessible to our direct control. Interventions that vary the relationship between brain activity and subjective experience will not only elucidate mechanisms that contribute to conscious experience, but help us understand why and how we are conscious at all. Here we outline an experimental program that will finally fix the hard problem of consciousness.

Significance

The causes of conscious sensation remains enigmatic. This is largely because from the perspectives of philosophy of mind and neuroscience there has been no agreement on how to experimentally address subjective conscious experiences. Here we argue the most productive approach will be an active, intervention-based program focused on altering the relationships between brain mechanisms and conscious experience in systematic ways.

Introduction

Consciousness depends on the brain. Neuroscientists have made significant progress in describing the relationship between the two. Yet what philosophers have dubbed the '*hard problem*' of consciousness suggests that this progress has a definite limit (1); (see SI§1**)**. After explaining all of the mechanisms that underlie conscious experience, there seems to be a further question about why the activity of these mechanisms is accompanied by any experience at all. As philosopher David Chalmers 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? (2, 201)

Why, in Nagel's (3) 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? The hard problem uniquely demands an explanation of *feels* rather than causes. This is why Chalmers (1) suggests that "…the hard problem seems to be a different sort of problem, requiring a different sort of solution" (1, pp 104).

We have not yet identified a brain mechanism for these subjective feelings, despite a lengthy search (4). We have identified detailed brain mechanisms for perception, cognition, and action, but leading philosophers of mind argue these mechanisms do not support or even require the rich inner mental life we experience as consciousness. Troublingly, we can imagine systems, both artificial or natural, that can be cognitively complex but have no subjective experiences (5). More troublingly, our subjective conscious experiences seem oddly arbitrary: why should middle C feel like *that* and not something else, and does middle C feel the same to everyone? Arbitrary relationships are anathema to scientific explanation.

Amongst philosophers, concerns raised by the hard problem have led to a variety of extreme views. On the one hand, there is growing interest in panpsychism, the position that consciousness is *everywhere*. If consciousness has an arbitrary relationship to the world, the line goes, why not treat it as a fundamental property akin to charge or mass (1)? At the other extreme, there is the assertion that debates about consciousness are some kind of pseudo-problem, because the phenomenon itself is a kind of illusion or philosophical confusion unsuitable for scientific study as such (6, 7).

Both routes limit the role of neuroscience, and the questions it may fruitfully ask. Experimentalists are caught in the middle. They often have the understandable impulse to avoid philosophical questions altogether, and to work on problems that admit of progress. We think this is a loss. The apparent mysteriousness of consciousness is another empirical fact in need of explanation. Experimental neuroscience can, and must, show us how to move beyond the hard problem. This paper details a scientific project that will allow just that.

Explaining Consciousness

The hard problem can be recast as a puzzle about scientific explanation. Neuroscience may discover many generalisations of the form "When *this* brain process occurs, people have an experience that feels like *that*." Yet these by themselves do not appear to admit of explanation. Why should one brain process give rise to the sensation of red and another the sensation of green, rather than vice-versa?

This contrasts with any other project in neuroscience. Elucidating the mechanisms of perceptual processes, say, explains why laws connecting discriminability to reaction times hold. In general, neuroscience might explain many facts about perceptual capacities: why this sort of brain process allows us to discriminate green things, to refer to them as 'green', or to judge that they are more similar to blue things than to red things. (These are the *easy problems.*) Yet when the explaining is done, it is still unclear why perceptual processes should have any *feel* to them, much less the particular feel that they have. This is all the more pressing given that many perceptual processes, including quite sophisticated ones, happen unconsciously.

We suggest that it's best to think of the hard problem as having two distinct but related aspects. The *objective* hard problem involves explaining the connection between a neural mechanism and a feeling. It is presently unclear how that works, and thus how it could be explained. The *subjective* hard problem reflects this difficulty from, as it were, the inside. It must address the apparent arbitrariness of the relationships between brains and felt experience. Why should mechanisms that perceive blue also come with a feeling of *blueness*?

Arguments for that arbitrariness tend to be philosophical in nature, relying primarily on thought experiments (see SI§2**)**. Scientists often respond that progress will remove the feeling of mystery that surrounds linking mechanisms to feelings. But the point of such thought experiments is to show that merely accumulating more of the same type of scientific evidence we currently have won't help.

We argue that moving beyond the hard problem will require a proper experimental, interventionist study of the hard problem of consciousness. Experimental manipulation is a key part of all natural science. Good explanations don't just tell you what will happen. They reveal part of the causal structure of the world: that is, they give you the *difference-makers* for the phenomenon you want to explain (8-12).

Most current work on consciousness aims for something less. The search for so-called Neural Correlates of Consciousness (NCCs) studies neural phenomena that are *associated* with conscious experience (13, 14). Subjects' conscious state can be manipulated by a variety of means. Yet the corresponding changes to brain activity are merely observed. The relationship between brain states and experiences (what we'll call *linking laws*) may thus be described, but are not explained (See SI§3). Further, work on NCCs is limited to conscious experiences that can be induced by external stimuli in neurotypical adults, which (one might worry) is a limited base from which to infer underlying mechanisms.

The core of interventionist science is finding causes that make a difference to their effects. Intervention is an active notion: one intervenes by keeping everything else fixed and varying the parameter of interest. Interventions distinguish between mere predictions and explanations. Barometer readings predict the storm but do not explain it. Why? Because we can't fiddle with the barometer to bring rain: the barometer reading is not a difference-maker for the storm (10). Smoking partially explains lung cancer rates, by contrast, because interventions which change the rate of smoking make a difference to cancer rates.

Direct brain intervention is complicated, of course. That is why NCC research is attractive. Yet we argue that direct brain intervention will be the key to solving the hard problem. The subjective and objective hard problems are two sides of the same coin; only direct brain intervention can handle both at once. Thus interventionist neuroscience can produce qualitatively better evidence about the mechanisms of consciousness.

Interventions that tell us something about both the objective and subjective aspects of the hard problem provide a way forward. We will detail what we can about that experimental project and how it ought to proceed. An important assumption is worth noting from the outset. We are physicalists. We take as a working assumption that the appearance of a hard problem is ultimately misleading. We might be wrong. If so, the project we propose would fail. The past successes of science in tackling similarly profound mysteries give us reason to believe that effort here will be repaid, but this is ultimately an empirical matter.

In the background is a simple vision worth keeping in mind. If it (mistakenly) seems as if there is a hard problem of consciousness, the appearance of a problem is ultimately a fact about our brains and how they are constructed. To make progress, then, neuroscientists must first find the neural basis for why our conscious experiences seem arbitrary and inaccessible. Having done so, we must then find suitable experimental interventions that eliminate the hard problem and render conscious experience intelligible. In other words, the hard problem doesn't need to be solved: it needs to be *fixed.*

Intervening on Consciousness

Unless otherwise specified, we assume that the interventions we describe occur in awake, neurotypical adults, and hence interventions will have both subjective and objective effects. Intervening on brains raises obvious practical and ethical complications. Nevertheless, there are already a variety of interventionist experiments that have been performed which illustrate how manipulations of the brain can profoundly change conscious experience.

Since Penfield's work (15, 16), we have known that direct electrical stimulation of the cortex can produce a wide variety of distinct experiences. Noninvasive stimulation (e.g. by TMS) can produce similar results, though with less specificity.

Interventions on particular experiences can happen in a second, indirect way. The same sensations can vary in how they feel given systematic alterations in broader underlying states. Consider, for example, phenomena such as 'morphine pain', where patients given a dose of morphine after acute injury will report that they continue to feel pain but no longer care about it. (Similar effects occur with a range of other dissociative drugs (17).) There is debate about whether this effect is due to a sensory-limbic dissociation (18) or to a general breakdown in processes of bodily ownership and concern (19), but in either case there appears to be alteration to the character of individual sensations by changes in the background conditions of experience.

Complex and subtle changes in the structural features of experience have also been reported. In a coda to *Awakenings,* Oliver Sacks (20) describes the subjective distortions of space and time that coexist with 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 also produce striking alterations in the perception of motion and time (21).

Finally, there are ways to alter experience even more profoundly. Work on anesthetics, or on patients with severe impairments of consciousness also looks more directly at the basic capacities that support subjective experience (22-24). Of course, many unsubtle and uninteresting interventions can cause unconsciousness. But as several authors have recently urged, the capacities which underlie consciousness, and hence the broader modes of variation possible, are probably numerous and heterogenous (25). There is no simple, well-ordered scale of 'degree' of awareness, for example; instead, there are numerous dimensions along which conscious experience as a whole might vary, and which plausibly link to the underlying functional capacities.

In the various interventions above, one should distinguish between interventions on the *contents* of conscious experience on the one hand and the *capacities* necessary for conscious experience on the other. Conscious contents are experiences like colors, sounds, shapes, or pains. The capacities necessary for conscious experience are (plausibly) 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 (26-29).

Capacities for experience are important, we suggest, because they often correspond to structural features of conscious experience. We experience visual and auditory sensations as occurring within a common, external space, for example. Yet while these structural features are the conditions for the possibility of conscious experience (30), they are arguably not themselves *objects* of experience. While we experience external sensations as occurring in a unified frame of reference, for example, we do not experience space as such, independent of the objects within it.

Roughly speaking, different brain regions appear responsible for capacities and contents of experience. Intervention on the cortex tends to change the contents of experience (15, 16). Work on the capacities for consciousness, by contrast, has more often focused on subcortical structures such as the midbrain and diencephalon. These are evolutionarily basal structures for vertebrates (27, 31). Further, the sorts of activities that the midbrain/diencephalon support are plausibly the sorts of activities that are conditions for consciousness. Bjorn Merker, for example, suggests that the integration of self-motion, exteroception, and internal valuation (supplanted by memory) are the key to bringing together a first-person perspective on the world (26, 27, 32, 33). Eva Jablonka similarly focuses on the role of integrative structures in supporting the capacity for unbounded associative learning (28, 29). We argue elsewhere that the development of integrative structures combining sensory percepts into a unified neural representation of the mobile animal within its environment is a form of major transition in neural evolution, which enabled a fundamental shift in behavioural capacity (26).

Solving the objective hard problem

The core of scientific explanation is showing why things are one way rather than another. To do so, one demonstrates how altering one aspect of the world makes a difference to the aspect you wish to explain. The same is true for consciousness. Different interventions one might make on brain states have the capability to explain different aspects of experience.

Figure 1 shows a progression of hypothetical interventions we might make on experience. To begin we might (1a) have evidence that activity in certain brain region *B* is associated with seeing a red object. Crucially, the claim is not that *B* alone is sufficient to give rise to a red experience. Rather, difference-makers always act against a large background of causal factors. Similarly, many brain regions might make a difference to the same aspect of experience, and interventions on the same brain region might have many effects on experience. There is a direct parallel here with genetics: most phenotypic traits are polygenic, a single gene can affect many traits, and genes only work in concert with a variety of other genes (34).

Explanation is always *contrastive*: we explain why the world is one way rather than another. Different contrast classes can require different explanations (8, 35, 36). One explains why the subject sees red *rather than* some other color, given that they have one of several possible regional patterns of brain activation. (This is, note, a simple example of a linking law relating a mechanism to experience (SI§3).) Other contrasts might invoke other brain regions: seeing saturated versus desaturated red, say, might be influenced by something other than *B*. Unlike versions of the NCC strategy, there is no assumption that there must be a unique region responsible for all aspects of an experience.

The linking law in 1a says more than that pattern B_1 is reliably correlated with seeing red. As shown in 1b, a direct intervention on B that changed its activation from B_1 to B_2 (and left everything else the same, as much as possible) would make the subject experience blue in the presence of the same stimulus. Capacities could be intervened upon in the same way. Figure 1c shows a single intervention that warps and distorts the perception of space, creating correlated changes across a variety of experiences.

Some interventions will be relatively crude. Many interventions will simply eliminate consciousness altogether by eliminating a necessary background condition (Figure 1d); intervention on the claustrum appears to act as a kind of on-off switch for experience (37). There may be many such interventions, and in general we are interested in interventions that give a more selective (38, 39) or systematic (9) handle on phenomena we care about. These are interventions 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.

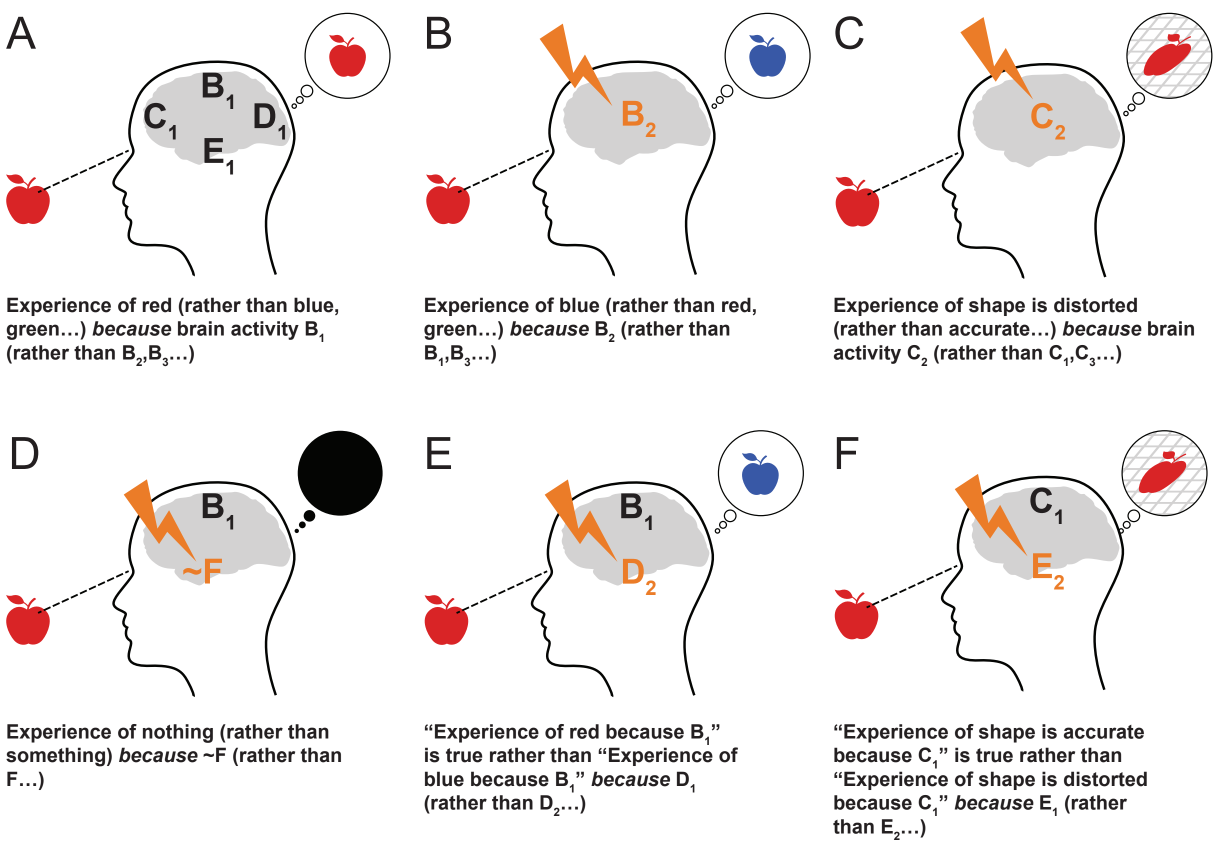

Figure 1: Different possible interventions of conscious states. (a) The simple case, explaining a token experience. B1 is a pattern of brain activation associated with seeing red, while C_1, D_1, E_1 etc are background conditions. The caption expresses a linking law. (b) An appropriate intervention on B will change the felt color 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) An intervention on a linking law. 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 sensation. (f) A similar intervention on structural capacities. Cases (a) and (b) are entirely compatible with property dualism (SI 1), whereas the remaining cases would be problematic.

Figure 1d also illustrates another important point. The intervention eliminates consciousness even in the presence of activity B_1 that gave rise to a red experience in ordinary circumstances. Another way to understand this is that the linking law in 1a holds only contingently; appropriate intervention changes the law from holding to not holding. In neuroscience, as in most disciplines, there are no exceptionless laws: laws are merely invariant across a range of circumstances (SI§3).

This is the key to explanation of linking laws themselves. Figures 1e and 1f show interventions which change laws linking mechanisms and experience in more systematic ways. In 1e, intervention on D changes the relationship between B and color experience. Activity in D gives us an explanation of why the linking law in 1a holds, *rather than* some other linking law. Figure 1f shows a similar intervention on linking laws regarding structural features of experience (see SI§4).

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. Systematic intervention on linking laws themselves is thus the way to explain them. And that is what the hard problem demands.

Solving the Subjective Hard problem

The story so far is the same that could be told about any complex scientific problem. But, the hard problem has an important subjective component as well. Recall that what we termed the *subjective hard problem* depends on the internal feeling that there is something arbitrary and mysterious about the connection between brain states and subjective experiences.

Interventions on brain states have an important subjective component as well: that is, by intervening on brains, we don't simply discover *that* certain experiences can be evoked, or that they depend on certain interventions. That is still third-person knowledge. The first-person, subjective experience of that intervention is important as well: it is one thing to read about (say) the experience of alien hand, and quite another to *feel* your fingers jump around under the influence of TMS to the motor cortex.

Yet we think there is more at stake than just proof of principle. If the hard problems have a solution, then the subjective *feeling* of a hard problem is mistaken (see SI§5).

Here we take inspiration from Hume, who noted regarding free will that

> There is a *false sensation* or *experience* even of the liberty of indifference; which is regarded as an argument for its real existence... We may imagine we feel a liberty within ourselves; but a spectator can commonly infer our actions from our motives and character; and even where he cannot, he concludes in general that he might, were he perfectly acquainted with every circumstance of our situation and temper, and the most secret springs of our complexion and disposition. Now this is the very essence of necessity... (40 2.3.2)

In other words, part of our subjective feeling that we are completely free and undetermined involves a certain lack of access to the causes of our own actions.

Consciousness works the same way. There is a view, tracing back at least to Leibniz, on which the apparent simplicity and arbitrariness of conscious states is merely an introspective confusion about a complex underlying state (41-43). 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" (44). There may be other sources of trouble as well, such as our relatively limited capacity for introspection and discussion of our conscious states compared to the richness of conscious experience itself (45).

In short, our subjective experience is underpinned by a great number of mechanisms to which we have no conscious access, and which 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 conscious states feel arbitrary: subjectively, they simply appear out of nowhere.

Intervention on consciousness changes this. By *feeling* how subjective experience is altered by altering brain activity, the feeling of arbitrariness should vanish. As with the objective hard problem, the most telling alterations are likely to be specific, systematic alterations of the capacities which underlie conscious experience itself.

To make the point vivid, consider Jackson's famous thought experiment (46). Mary is a scientist who has been raised in a black-and-white room. She is in possession of a completed neuroscience, and so knows all there is to know about color perception. Yet she's never seen red. Intuitively, Jackson claims, when she steps outside of the room she'll learn something new: what red looks like to other people. Since she already knew all of the relevant neuroscience, her gain in knowledge shows that experience outstrips neuroscience. There are many philosophical responses to this argument (47). But we think an interventionist perspective allows instead for a useful tweak on the setup.

Consider Mary's alien counterpart, Proud Mary. Her neurobiology is very different to ours. She has far more extensive sets of feedback loops in the brain. We have the ability to (for example) rapidly and unconsciously 'compile' commands in natural language and turn them into motor commands (48). Proud Mary can do much more: with minimal effort, she can alter the functioning of any circuit in the brain (including the circuits responsible for that very ability) in accordance with any description she'd like. Reading about how red works just gives her the experience of red. Indeed, her experience can be arbitrarily altered with minimal effort of will.

Though we can conceive of Proud Mary, it's difficult to know what it would be like to be her. (Indeed, it might be like almost anything.) One thing we *do* think, however, is that Proud Mary would not be particularly impressed by the hard problem. For she is not a passive bystander to her experiences. The move from textbook to experience would seem to her no more mysterious than the move from a score to a song would seem to a skilled pianist. Further, just as a skilled musician can transpose and improvise on the fly, so too could Proud Mary alter broad features of her subjective experience.

It seems unlikely to us that Proud Mary would recognize a distinction between the hard and easy problems of consciousness. Hence the subjective hard problem depends in part on contingent features of our own neurological makeup. If the basis of the subjective hard problem is neurobiological, then the solution also lies in understanding that neurobiology and gaining the tools to manipulate it. Crucially, what we will gain thereby is not just methods for manipulation, but for *self*-manipulation.

Making progress

To make progress in fixing the objective and subjective hard problems we need techniques for specific and selective interventions on conscious experience. Such interventions are difficult to come by in humans; at present they are either pharmacological, invasive, or non-invasive. We have noted some possibilities above, and we consider each in turn.

Pharmaceutical intervention is the most familiar and accessible way to intervene upon consciousness. There has been a recent revival of interest in psychedelics given their promising results in treating conditions like PTSD. That said, we caution against repeating the mistakes of the past. In particular, we think it is worth being wary of returning to the uncritical pharmacological investigations that were popular in the past (49, 50). 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 (51). This is an old idea, embodied in the etymology of 'psychedelic' itself. We are skeptical. 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. For example, reports suggest that low doses of diisopropyltryptamine (DiPT) have effects primarily limited to nonlinear distortions of audition (52). Limited and well-defined phenomena may also be fruitfully investigated, as for example in work done using LSD to investigate the central mechanisms of binocular rivalry (53).

Invasive interventions involving direct electrical stimulation of the brain have been important to understand conscious function (16). Invasive work presents obvious ethical and practical concerns, and so is typically only 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 posterior cingulate/precuneus can produce more profound alterations in global experience (54-56). This would be consonant with these regions' purported role in consciousness and mediating cortical-subcortical interactions (57, 58).

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 (59, 60) or induction (61) of psychiatric conditions such as obsessive-compulsive disorder and depression. Thalamic DBS has also led to promising improvements in minimally conscious patients (62). The variety of possible stimulation parameters, and the variability of results between microstimulation and direct electrical stimulation (63) suggests a fruitful experimental program 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 think that this ought to be more routinely and systematically collected.

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 (sensu 64); the aim being to supplement or replace memory, or even add new information representations (65-67). We suspect that these particular forms of brain-machine interface will not help us much with the hard problem. Working out how information is represented in the brain remains an easy problem. On the other hand, if (say) an artificially implanted memory somehow *felt* distinctive, we might be able to learn more about the hard problem by using this as a contrast case.

Finally, noninvasive 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 (68). 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 (69, 70). tES has had problems showing specificity and replicability, but recent techniques using EEG/MEG to guide stimulation timing (71) 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.

Much of the work on developing new forms of brain-machine interface is currently happening with animals (65-67). 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 (31), and of neuron functions across most animal phyla (72), 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 don’t know what neural circuits are necessary to support conscious experience (24, 73). Further, solving the subjective hard problem ultimately requires self-intervention, so animal models can only ever do part of the job.

That said, we envision research with animal models to play 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 (24, 26).

7. Conclusion: Fixing the hard problem

We have outlined an ambitious program for solving the hard problem. The hard problem, in both its aspects, stems from our own limitations. Having identified these, neuroscientists must fix those shortcomings. This requires direct intervention, and a mix of third-person and first-person techniques.

Our proposal may strike authors from certain philosophical traditions as odd. Surely the hard problem was about explaining *consciousness*, not why the laws connecting brain to experience have this feature rather than that. Furthermore, contrastive explanation is by its nature pluralist: there will be not one grand explanation but many interlinking explanations.

The history of science provides numerous 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. In the 18th century, there was a grand philosophical challenge to explain *Life*. Considered as such, little progress could be made.

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 this process kept blood pH within reasonable limits while that process cleared carbon dioxide rather than letting it accumulate, and so on. 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 (74). 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, 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, but no more mysterious, than life.

Finally, and crucially, we emphasise that this is a research program which is fundamentally falsifiable. That is, we might find that there are no systematic ways to intervene on linking laws: they are in fact like the brute laws of fundamental physics. Were it 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 that will be required are only dimly understood at present. We can do a lot of science, and indeed will have to, before embarking on such a project. Rather, the claim is that until we reach this final step, 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 argued that we do not know what it is like to be a bat, his point was not to argue against physicalism (3, pp 447). 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 tells us, his argument should be seen as "a challenge to form new concepts and devise a new method" of approaching experience (3, pp 449). We agree, we just think that a more direct approach is required. The hard problem will not be solved by philosophical discussion of positions relative to the problem alone, be they illusionist (S1§5), dualist, physicalist or panpsychist. The difference between our proposal and the present philosophical impasse is akin to the difference between Freudian psychoanalysis and modern pharmaceutical approaches to mental illness. Increased understanding is important, and conceptual change is inevitable. But there is no 'talking cure' for the hard problem: some degree of direct intervention will be necessary.

References

1. Chalmers DJ (2003) Consciousness and its place in nature. *Blackwell guide to the philosophy of mind*, eds Stich SP & Warfield TA (Blackwell), pp 102-142.
2. Chalmers D (1996) *The Conscious Mind: In Search of a Fundamental Theory* (Oxford University Press, New York).
3. Nagel T (1974) What is it like to be a bat? *The Philosophical Review* 83(4):435-450.
4. Snaprud P (2018) The consciousness wager. in *New Scientist*, pp 28-31.
5. Chalmers DJ (1996) *The Conscious Mind: In Search of a Fundamental Theory* (Oxford University Press, New York).
6. Dennett DC (1991) *Consciousness Explained* (Little, Brown, & co., Boston).

7. Irvine E (2013) *Consciousness as a scientific concept: A philosophy of science perspective* (Springer, New York).
8. Woodward J (2003) *Making Things Happen* (Oxford University Press, New York).
9. Klein C (2017) Brain Regions as Difference-Makers. *Philosophical Psychology* 30((1-2)):1-20.
10. Craver CF (2007) *Explaining the brain* (Oxford University Press, New York).
11. Campbell J (An interventionist approach to causation in psychology. Causal Learning: Psychology, Philosophy and Computation, pp 58-66.
12. Miracchi L (in press) Generative explanation in cognitive science and the hard problem of consciousness. *Philosophical Perspectives*.
13. Crick F & Koch C (1990) Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences* 2:263-275.
14. Chalmers DJ (1998) On the Search for the Neural Correlate of Consciousness. *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates* 2:219.
15. Penfield W & Jasper H (1954) *Epilepsy and the functional anatomy of the brain* (J.&A. Churchill, Ltd, London) pp 122-126.
16. Penfield W & Rasmussen T (1950) *The cerebral cortex of man: a clinical study of localization of function.* (Macmillan, New York).
17. Keats AS & Beecher HK (1950) Pain relief with hypnotic doses of barbiturates and a hypothesis. *Journal of Pharmacology and Experimental Therapeutics* 100(1):1-13.
18. Grahek N (2007) *Feeling pain and being in pain* (MIT Press, Cambridge).
19. Klein C (2015) What pain asymbolia really shows. *Mind* 124(494):493-516.
20. Sacks O (Parkinsonian Space and Time. Awakenings, pp 339-349.
21. Wolfe TR & Caravati EM (1995) Massive dextromethorphan ingestion and abuse. *The American journal of emergency medicine* 13(2):174-176.
22. Alkire MT, Hudetz AG, & Tononi G (2008) Consciousness and anesthesia. *Science* 322(5903):876-880.
23. Mashour GA & Alkire MT (2013) Consciousness, anesthesia, and the thalamocortical system. *Anesthesiology* 118(1):13-15.
24. Klein C & Barron AB (2016) Insects have the capacity for subjective experience. *Animal Sentience*:100.
25. Bayne T, Hohwy J, & Owen AM (2016) Are there levels of consciousness? *Trends Cogn Sci* 20(6):405-413.
26. Barron AB & Klein C (2016) What insects can tell us about the origins of consciousness. *Proc Nat Acad Sci USA* 113(18):4900-4908.
27. Merker B (2007) Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences* 30:63-81.
28. Ginsburg S & Jablonka E (2007) The transition to experiencing: the evolution of associative learning based on feelings. *Biol Theor* 2(3):231-243.
29. Ginsburg S & Jablonka E (2010) The evolution of associative learning: a factor in the cambrian explosion. *J Theor Biol* 266(1):11-20.
30. Kant I (1999) *Critique of Pure Reason: The Cambridge Edition of the Works of Immanuel Kant* (Cambridge University Press, Cambridge).
31. Striedter GF (2005) *Principles of brain evolution* (Sinauer Associates, Sunderland, MA).
32. Merker B (2005) The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition* 14:89-114.

33. Merker B (2013) The efference cascade, consciousness, and its self: naturalizing the first person pivot of action control. *Frontiers in Psychology* 4:501.
34. Sterelny K & Kitcher P (1988) The return of the gene. *The Journal of Philosophy* 85(7):339-361.
35. Hitchcock CR (1996) The Role Of Contrast In Causal And Explanatory Claims. *Synthese* 107:395-419.
36. van Fraassen BC (1980) *The Scientific Image* (Oxford University Press, New York).
37. Koubeissi MZ, Bartolomei F, Beltagy A, & Picard F (2014) Electrical stimulation of a small brain area reversibly disrupts consciousness. *Epilepsy & Behavior* 37:32-35.
38. Woodward Hopf F & Bonci A (2010) Dnmt3a: addiction's molecular forget-me-not? *Nature Neuroscience* 13(9):1041-1043.
39. Griffiths PE*, et al.* (2015) Measuring Causal Specificity. *Philosophy of Science* 82(4):529-555.
40. Hume D (2000) *A Treatise of Human Nature* (Oxford University Press, Oxford).
41. Hilbert DR (1987) *Color and color perception: A study in anthropocentric realism* (Center for the Study of Language and Information, Stanford).
42. Armstrong D (1997) "What is consciousness?" *The nature of consciousness: Philosophical debates*, ed N Block, OJ Flanagan, and G Güzeldere. Cambridge: MIT Press, pp 721-728.
43. Pettit P (2003) Looks as powers. *Philosophical Issues* 13(1):221-252.
44. Lashley KS, Beach FA, Hebb DO, Morgan CT, & Nissen HW (1960) The neuropsychology of Lashley: selected papers of KS Lashley.
45. Block N (2011) Perceptual consciousness overflows cognitive access. *Trends Cogn Sci* 15(12):567-575.
46. Jackson F (1982) Epiphenomenal qualia. *The Philosophical Quarterly* 32:127-136.
47. Ludlow P, Nagasawa Y, & Stoljar D eds (2004) *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument* (MIT Press, Cambridge Mass.), p 484.
48. Hamburger H & Crain S (1984) Acquisition of cognitive compiling. *Cognition* 17(2):85-136.
49. Jay M (2009) The atmosphere of heaven: the 1799 nitrous oxide researches reconsidered. *Notes and Records*:rsnr20090006.
50. Lattin D (2010) *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* (Harper Collins).
51. Letheby C (2015) The philosophy of psychedelic transformation. *Journal of Consciousness Studies* 22(9-10):170-193.
52. Shulgin A (2000) *Tryptamines I have known and loved* (Mind Books, Berkeley).
53. Carter OL & Pettigrew JD (2003) A common oscillator for perceptual rivalries? *Perception* 32(3):295-305.
54. Balestrini S*, et al.* (2015) Multimodal responses induced by cortical stimulation of the parietal lobe: a stereo-electroencephalography study. *Brain* 138(9):2596-2607.
55. Herbet G*, et al.* (2014) Disrupting posterior cingulate connectivity disconnects consciousness from the external environment. *Neuropsychologia* 56:239-244.
56. Herbet G, Lafargue G, & Duffau H (2015) The dorsal cingulate cortex as a critical gateway in the network supporting conscious awareness. *Brain* 139(4):e23.

57. Vogt BA & Laureys S (2005) Posterior Cingulate, Precuneal & Retrosplenial Cortices: Cytology & Components of the Neural Network Correlates of Consciousnes. *Progress in Brain Research* 150:205-217.
58. Cavanna AE & Trimble MR (2006) The precuneus: A review of its functional anatomy and behavioural correlates. *Brain* 129(3):564-583.
59. Krack P, Hariz MI, Baunez C, Guridi J, & Obeso JA (2010) Deep brain stimulation: from neurology to psychiatry? *Trends in neurosciences* 33(10):474-484.
60. Lyons MK (2011) Deep brain stimulation: current and future clinical applications. in *Mayo Clinic Proceedings*, pp 662-672.
61. Bejjani B-P*, et al.* (1999) Transient acute depression induced by high-frequency deep-brain stimulation. *New England Journal of Medicine* 340(19):1476-1480.
62. Schiff ND*, et al.* (2007) Behavioural improvements with thalamic stimulation after severe traumatic brain injury. *Nature* 448(7153):600.
63. Vincent M*, et al.* (2016) The difference between electrical microstimulation and direct electrical stimulation-towards new opportunities for innovative functional brain mapping? *Reviews in the Neurosciences* 27(3):231-258.
64. Clark A (1995) Moving Minds: Situating Content in the Service of Real-Time Success. *Philosophical Perspectives* 9:89-104.
65. Berger TW*, et al.* (2011) A cortical neural prosthesis for restoring and enhancing memory. *Journal of Neural Engineering* 8(4):046017.
66. Deadwyler SA*, et al.* (2013) Donor/recipient enhancement of memory in rat hippocampus. *Frontiers in Systems Neuroscience* 1:120.
67. Deadwyler SA*, et al.* (2017) A cognitive prosthesis for memory facilitation by closed-loop functional ensemble stimulation of hippocampal neurons in primate brain. *Experimental Neurology* 287(4):452-460.
68. Thibaut A, Bruno M-A, Ledoux D, Demertzi A, & Laureys S (2014) tDCS in patients with disorders of consciousness Sham-controlled randomized double-blind study. *Neurology* 82(13):1112-1118.
69. Filmer HL, Dux PE, & Mattingley JB (2014) Applications of transcranial direct current stimulation for understanding brain function. *Trends in neurosciences* 37(12):742-753.
70. Tavakoli AV & Yun K (2017) Transcranial alternating current stimulation (tACS) mechanisms and protocols. *Frontiers in cellular neuroscience* 11:214.
71. Thut G*, et al.* (2017) Guiding transcranial brain stimulation by EEG/MEG to interact with ongoing brain activity and associated functions: a position paper. *Clinical Neurophysiology* 128(5):843-857.
72. Kristan WB (2016) Early evolution of neurons. *Curr. Biol.* 26:R937–R980.
73. Klein C & Barron AB (2016) Insect consciousness: commitments, conflicts and consequences. *Animal Sentience* 9:21.
74. Machery E (2012) Why I stopped worrying about the definition of life... and why you should as well. *Synthese* 185:145-164.

For PNAS Perspectives

Supplementary Material for

“How Experimental Neuroscientists Can Fix the Hard Problem of Consciousness”

Colin Klein and Andrew B. Barron

§ 1:

Philosophers generally distinguish the *easy problems* of consciousness from the *hard problem* (1). Easy problems involve detailing the causal mechanisms which underlie perceptual discrimination, episodic memory recall, and so on—the sort of thing that cognitive neuroscience already does well. The hard problem involves explaining why any of these causal mechanisms feel like something.

Within this literature, a *physicalist* is someone who believes that all facts, including facts about subjective experience and their basis, ultimately depend upon ordinary facts about physical objects and their properties. A *dualist* thinks that there is a category of being over and above the physical and its properties. The most familiar dualist, Descartes, was a *substance dualist:* very roughly, he thought that there were two kinds of stuff in the world, mental and physical. Most contemporary dualists are *property dualists*: they believe there is only physical stuff, but at least some of it has distinct, non-physical properties. (Drawing the boundary between physical and nonphysical can itself be tricky; see (2).) Note crucially for what follows that the phenomenal properties are *epiphenomenal* on most accounts: while they are caused, they cause nothing further.

§2:

Part of the bite of the hard problem comes from the fact that the link between subjective experience and mechanisms seems fundamentally arbitrary. Arguments for the existence of the hard problem, to a large extent, rest on philosophical thought experiments designed to pull out this feeling.

Leibniz’s Mill is an indicative early example. As the philosopher Gottfried Leibniz put it:

> …supposing that there were a mechanism so constructed as to think, feel and have perception, we might enter it as into a mill. And this granted, we should only find on visiting it, pieces which push one against another, but never anything by which to explain a perception (3)

Several thought experiments are particularly important for understanding the modern debate. Thomas Nagel (4) argued that no matter what we learn about brains, we are fundamentally limited in our ability to understand what it is like to be a creature like a bat. Frank Jackson (5) has us imagine Mary, a neuroscientist who knows all there is to know about the brain but who is raised in a black-and-white room. When released from her monochrome lab, she learns what it is like for others to see red, a fact she could not know from learning about brain mechanisms alone.

The argument from spectrum inversion imagines swapping the properties of experience with no effect on behaviour (6). We have a sensation of blue when we see the sky, and of yellow when we see a daffodil. Why couldn't those be swapped? It seems like the particular sensations don't matter; they are a kind of 'mental paint' (7) that could have been chosen differently with the same effects. David Foster Wallace imagines a budding philosopher who

> ...is struck by the ghastly possibility that, e.g., what he sees as the color green and what other people call "the color green" may in fact not be the same color or experience at all: the fact that both he and someone else call Pebble Beach's fairways green an a stoplight's GO signal green appears to guarantee only that there is a similar consistency in their color experiences of fairways and GO lights, not that the actual subjective quality of those color experiences is the same; it could be that what [he] experiences as green everyone else actually experiences as blue, and what we "mean" by the word *blue* is what he "means" by *green*, etc... (8, fn 23)

This thought experiment highlights that it doesn't seem very important *which* subjective experiences we have: so long as they were consistent over an individual's life, blue and green could be swapped with no effect.

Capping it off, David Chalmers claims that we can conceive of philosophical zombies: things that look and act exactly like people who are capable of doing everything we do but which lack subjective experience altogether (1, 9). If so, then subjective experience could be completely *eliminated* without effect on behaviour.

There is a tendency for experimental scientists to be dismissive of conceivability arguments. We think this view is unconstructive, and has contributed to the failure of experimental neuroscience to properly engage with the hard problem. Conceivability arguments, in the guise of thought experiments, are an important scientific tool with a venerable history (10). In his autobiography, for example, Einstein recalls

> ...a paradox upon which I had already hit at the age of sixteen: if I pursue a beam of light with the velocity c (velocity of light in a vacuum), I should observe such a beam of light as a spatially oscillatory electromagnetic field at rest. However, there seems to be no such thing, whether on the basis of experience or according to Maxwell's equations. (11)

Conceivability arguments seem especially suited to learning about consciousness. It feels as if all you need to know about the *sensation* of color (as opposed to its causes) is revealed to you just by having the sensation itself (12, 13). Of course, we ultimately deny this. And one might reasonably worry that conceivability is a poor guide to learning things about the mind, particularly given the limitations of introspection (14). But in general, if one thinks that a conceivability argument is misguided, it is worth having a story about why it is that we can end up confused about the conceivable.

§3:

"Linking law" is a deliberate terminological compromise. As we understand them, linking laws have two important features. First, like traditional psychophysical bridge laws, they link up physical and phenomenal states. Like all laws, they are counterfactual-supporting and change-relating, not simply descriptive. If an experience of blue corresponds to brain state B1 and an experience of red to B2, then changing from B1 to B2 should change the experience.

However, we also intend linking laws to be understood on the model of Woodward's invariant generalisations (15). They need hold only under a limited range of circumstances and interventions (as is, in fact, the case for nearly all invariant generalisations, special science or otherwise). This is a metaphysical claim, note, not a claim about explanation: in accord with contemporary thinking on modelling in science, it's quite possible that the best way to discuss laws themselves is via a model-based strategy.

Linking laws in this sense are importantly different from standard *ceteris paribus* (CP) laws familiar to philosophers of mind. On standard accounts of CP laws, ordinary statements of CP laws are shorthand for ones with the antecedent appropriately restricted by a clause which, when incorporated, makes the law in fact strict (16-19).

There are various problems with CP laws so understood, and the move to invariant generalisations is part of a move away from the idea that laws must be strict to be explanatory (and therefore special science laws). A promising alternative is to treat linking laws as externally restricted, perhaps via tacit quantifier domain restriction (20). In addition to the modern interventionists, Hempel endorses something like it for the role of CP-clauses (21). Lange has argued for what we take to be a similar view (22, 23). Klee also endorses a similar position, attributing it to Lakatos (24).

§4:

Most work on interventions focuses on the relationship between event types. However, the same logic can be generalised to the explanation of invariant generalisations themselves, including to linking laws. That is, we can explain some linking law *L* by showing facts which could be varied in order to vary *L*. This is not available under traditional accounts of explanations of laws, because laws cannot vary; as per SI 1§3, even ceteris paribus laws are understood as exceptionless when properly hedged.

By contrast, explanatory generalisations needn't be invariant over any intervention and in any background. Indeed, one advantage of the interventionist account was precisely that it admitted as explanatory generalisations which might be more or less fragile.
That provides a natural extension to explanations of generalisations themselves. This has been rarely discussed (though see (25) Chapter 4), but it is at least implicit in some remarks by Woodward. Consider, for example, Woodward's discussion of the acceleration of a wooden block down an incline of slope Φ (2003: 13-14). After deriving the standard equation:

$$a = g \sin \Phi - \mu_K g \cos \Phi$$

Woodward notes that

> ... the above explanation allows us to see how changing the values of various variables ... would result in changes in what it is that we are trying to explain: the acceleration of the block. For example, we see ... how changing the angle of elevation of the plane will change the acceleration of the block: as the angle is increased, the acceleration will be greater. Similarly, the above explanation shows us how, if we were to move the apparatus to a stronger or weaker gravitational field (i.e., if we were to change the value of the "constant" *g*), the acceleration of the block would change. We can also see, by way of contrast, that changing the value of *m* will make no difference to the motion of the block (15).

Note that the explanatory information cited by Woodward is not relevant merely to a particular kind of block or ramp, but rather to the general relationship between blocks and ramps: it says that this relationship would be invariant under a certain class of manipulations and vary in systematic ways given others.

Figure 1 in the main text shows the fundamental parallelism between this kind of explanation and explanation of token events. The contrast class switches from values that particular variables might take to different ways that the linking law might be. These changes might be in particular parameters, in the general functional form of the equation, or something more complex. Any axis along which laws might vary is fair game.

This is far more information than given by CP clauses, even were CP clauses fully elaborated. CP-clauses tell us only that a relationship breaks down under certain conditions. The explanatory force of the information that Woodward cites, by contrast, tells us how a generalisation might *vary* given variations in other conditions. Just as relationships between event types can be more systematic and specific, so too (we claim) for relationships between conditions and generalisations.

One might object that the explanatory information provided merely serves to *parameterise* a given generalisation. Indeed, in the case that Woodward provides, a variable representing the gravitational force g is explicitly contained within the explained generalisation itself. The problem of variable choice is a complex one (26). However, there is no reason why factors with systematic influence must always be represented within a law itself. For one, explanatory generalisations need not contain within themselves all relevant factors—indeed, the denial of this is, again, the difference between the DN and the interventionist model. Finally, one might want to separate out distinct factors which potentially make a difference to a wide variety of different generalisation. So for example, while the frictional force μ_K is specific to individual setups, a change in the gravitational force g would affect a huge number of similar systems, in similar ways. We might therefore wish to treat it as an external parameter in order to capture its role across a variety of systems. This is even more plausible if our explanatory concerns are ones where g itself does not change in the ordinary course of things, in which case it might simply be replaced by a constant.

One might also object that the sort of explanation that is proper to neuroscience is, broadly speaking, mechanistic (27). Indeed, one of Chalmers' arguments for property dualism relies on the alleged incompatibility between raw feels and causal-functional explanation (1). Of course, we think that mechanistic explanation will play a critical role in explaining consciousness. We also think that experiment will ultimately fix the intuition that there is a conflict between mechanistic explanation and the nature of consciousness. But there is a broader point at stake as well. Mechanistic explanation (at least *sensu* Craver) is a species of interventionist explanation. We think that the latter is the more general notion, and that insofar as mechanistic accounts explain it is because they have hooked onto the sort of causal structure that is relevant for interventionist explanation. This also allows for (e.g.) mathematical explanation in the absence of causal structure (28).

SI 5:

As per the introduction, we believe that there is at least an appearance of a hard problem. It is important to distinguish our position from the philosophical view that Keith Frankish calls *illusionism* (29). This view, most often associated with the work of Daniel Dennett, holds that the hard problem is based on the illusion that there is phenomenal experience at all.

We are sympathetic to these positions, but our view is importantly different. First, while we think that the hard problem is based upon an illusion, we don't think that subjective experience itself is an illusion. This may be just a semantic point: the line between illusion and error is difficult to draw.

More broadly, our position is closer to Chalmers' Type C materialism (30), which holds that the hard problem will be solved by conceptual change brought by scientific advances. Illusionism is a version of Type A materialism, on which scientific advances will show the inappropriateness of the concept itself.

Second, and more importantly, we differ on what ought to be *done* about the hard problem. Indeed, both illusionism and type C materialism both seem to imply that the deep problem is one with our language or our conceptual scheme: learn enough, or do enough philosophy, and the hard problem goes away. The illusion might persist (in the same way that visual illusions persist even when you know the trick), but the sense of mystery will vanish. By contrast, we think that the fundamental problem is a causal one, and which requires a scientific solution in a particularly deep sense.

1. Chalmers DJ (1996) *The Conscious Mind: In Search of a Fundamental Theory* (Oxford University Press, New York).
2. Wilson J (2006) On characterizing the physical. *Philosophical Studies* 131(1):61-99.
3. Leibniz G (1714/1997) *Monadology*; trans Latta R & Rutherford D.
4. Nagel T (1974) What is it like to be a bat? *The Philosophical Review* 83(4):435-450.
5. Jackson F (1982) Epiphenomenal qualia. *The Philosophical Quarterly* 32:127-136.
6. Block N (Troubles with Functionalism. Perception and Cognition, Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science, pp 261-325.
7. Block N (1996) Mental paint and mental latex. *Philosophical issues* 7:19-49.
8. Wallace DF (2001) Tense Present Democracy, English, and the Wars over Usage. *Harper's Magazine*.
9. Hart WD, Hart W, & others (1988) *The engines of the soul* (CUP Archive).
10. Gendler TS (2000) *Thought experiment: On the powers and limits of imaginary cases* (Routledge, New York).
11. Einstein A (1949) Notes for an Autobiography. *The Saturday Review of LIterature*:9-12.
12. Kripke SA (1981) *Naming and necessity* (Wiley-Blackwell, New York).
13. Johnston M (1992) How to Speak of the Colors. *Philosophical studies* 68(3):221-263.
14. Schwitzgebel E (2008) The unreliability of naive introspection. *Philosophical Review* 117(2):245-273.
15. Woodward J (2003) *Making Things Happen* (Oxford University Press, New York).
16. Fodor J (1991) You Can Fool Some of the People All of the Time, Everything Else Being Equal; Hedged Laws and Psychological Explanations. *Mind* N.S. 100:19-34.
17. Pietroski P & Rey G (1995) When Other things aren't equal: Saving Ceteris Paribus Laws from Vacuity. *The British Journal for the Philosophy of Science* 46(1).
18. Schurz G (2001) Pietroski and Rey on Ceteris Paribus Laws. *The British Journal for the Philosophy of Science* 52:359-370.
19. Earman J & Roberts J (1999) C*eteris Paribus,* there is no problem of provisos. *Synthese* 118:439-478.
20. Stanley J & Szabó ZG (2000) On Quantifier Domain Restriction. *Mind and Language* 15(2 and 3):219-261.
21. Hempel CG (1988) Provisoes: A Problem Concerning the Inferential Function of Scientific Theories. *Erkenntnis* 28:147-164.
22. Lange M (1993) Natural Laws and the Problem of Provisos. *Erkenntnis* 38:233-248.
23. Lange M (2002) Who's afraid of C*eteris-Paribus* Laws? Or: How \mboxI Learned to Stop Worrying and Love Them. *Erkenntnis* 57:407-423.
24. Klee R (1992) Anomalous Monism, Ceteris Paribus, and Psychological Explanation. *The British Journal for the Philosophy of Science* 43(3):389-403.
25. Gijsbers V (2011) Explanation and determination. (Leiden University).
26. Woodward J (2016) The problem of variable choice. *Synthese* 193(4):1047-1072.
27. Craver CF (2007) *Explaining the brain* (Oxford University Press, New York).
28. Gijsbers V (2017) A Quasi-Interventionist Theory Of Mathematical Explanation. *Logique & Analyse* 237:47-66.
29. Frankish K (2016) Illusionism as a theory of consciousness. *Journal of Consciousness Studies* 23(11-12):11-39.
30. Chalmers DJ (2003) Consciousness and its place in nature. *Blackwell guide to the philosophy of mind*, eds Stich SP & Warfield TA (Blackwell), pp 102-142.