PhiMiSci
Philosophy and the Mind Sciences

Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs)

Colin Klein[a] (colin.klein@anu.edu.au)
Jakob Hohwy[b] (Jakob.Hohwy@monash.edu)
Tim Bayne[b] (Timothy.Bayne@monash.edu)

Abstract
The science of consciousness is currently structured around the search for the neural correlates of consciousness (NCCs). One of the alleged advantages of the NCC framework is its metaphysical neutrality—the fact that it begs no contested questions with respect to debates about the fundamental nature of consciousness. Here, we argue that even if the NCC framework is metaphysically neutral, it is structurally committed, for it presupposes a certain model—what we call the Lite-Brite model—of consciousness. This represents a serious liability for the NCC framework, for the plausibility of the Lite-Brite model is very much an open question, and the science of consciousness would be better served by a framework that does not presuppose it. Drawing on interventionist ideas in the philosophy of science, we suggest that the Difference-Maker framework can provide just such an alternative. Instead of searching for the neural correlates of consciousness (NCCs), we ought to be searching for the difference makers of consciousness (DMCs).

Keywords
Consciousness · Difference-making· Explanation · Intervention · Neural correlates · NCCs

This article is part of a special issue on "The Neural Correlates of Consciousness", edited by Sascha Benjamin Fink.

[a]The Australian National University
[b]Monash University

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60

©The author(s). https://philosophymindscience.org ISSN: 2699-0369

1 Introduction

Since the revival of neuroscientific interest in consciousness in the early 1990s, consciousness science has focused on the search for the neural correlates of consciousness (NCCs) (Koch et al., 2016; Metzinger, 2000). Although there is much debate about what and where the NCCs are (e.g. Boly et al., 2017; Odegaard et al., 2017), there has been relatively little debate about whether the science of consciousness ought to be framed in terms of the search for the NCCs (although see Noë & Thompson, 2004; Neisser, 2012).

This paper challenges that consensus, arguing that the NCC framework presupposes a particular – and, we suggest, implausible – conception of the structure of consciousness. We call the conception in question *the Lite-Brite model of consciousness*. The Lite-Brite model and the NCC framework are mutually reinforcing: adherence to the NCC framework as the only game in town supports the assumption that the Lite-Brite model must be right, and that in turn underpins the sense that the NCC framework is appropriate.

In the first half of this paper we argue that the science of consciousness would be better served by adopting a framework which is neutral between competing conceptions of the structure of consciousness. In the second half of the paper we turn to the task of developing such a framework. Drawing on interventionist accounts of explanation (Craver & Kaplan, 2011; Cummins, 2000; Pearl, 2000; Woodward, 2003), we suggest that the goals of consciousness science would be best-served by focusing on difference-making relations. In a nutshell, the science of consciousness should replace the search for the neural correlates of consciousness with the search for the difference-makers of consciousness (DMCs).

2 The NCC framework

At the heart of the NCC framework is the very notion of an NCC. In a representative recent statement, Koch et al (2016, p. 308) define an NCC as "… the minimum neuronal mechanisms jointly sufficient for any one specific conscious percept." This requires some unpacking; we will do our critique in the course of that unpacking. We note from the outset that the notion of a neural correlate of consciousness is a contested one, and there is some variation in how different research groups understand the notion. That said, there is enough of a consensus about what an NCC is for us to intelligibly talk about 'the NCC framework' – that is, the idea that consciousness science should be structured around the search for NCCs.

First, although Koch et al. refer here to 'neural mechanisms', there are few discussions of mechanisms or mechanistic explanation as such within the NCC literature. Indeed, the question of how exactly to characterize the N's in the NCCs has been relatively neglected. Early work on NCCs tended towards a simple localizationist approach in which Ns were categorized in terms of brain regions. For example, V5/MT has often been described as the NCC for motion perception (Block,

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60


©The author(s). https://philosophymindscience.org ISSN: 2699-0369

2005). More recent work suggests a range of ways of categorizing N-type states. In other areas of neuroscience, for example, authors have emphasized the explanatory importance of structural and dynamic networks (Sporns, 2011).

Second, although Koch et al. refer to 'one specific conscious percept', the NCC framework has not been restricted to percepts (or even to the contents of consciousness more generally). The NCCs that are specific to particular conscious contents are typically called content-NCCs. In addition to content-NCCs, theorists have also targeted state-NCCs – that is, the neural correlates of global, non-specific, conscious states, such as those that are associated with alert wakefulness, REM-dreaming, light sedation and the minimally conscious state (Bayne et al., 2016; Laureys, 2005). A third type of conscious state that has been brought within the scope of the NCC framework is the generic state of simply being conscious (Chalmers, 2000). An NCC which functions as a minimal neuronal substrate for being conscious is a generic-NCC.[1]

But it is the third aspect of the notion of an NCC that is perhaps most noteworthy: the idea that a neural state *N* qualifies as the NCC for a particular conscious state *C* if and only if *N* is 'minimally sufficient' for *C*. *N* is minimally sufficient for *C* if and only if the existence of *N* necessitates the existence of *C* (that's the sufficiency bit), and no proper part of *N* necessitates *C* (that's the minimal bit). The motivation for characterizing NCCs in terms of *minimal* sufficiency is clear: one wants to isolate only those neural features that are *directly* implicated in consciousness. From the perspective of the NCC framework, neural states that are causally upstream or downstream of consciousness ought to be screened off as confounds (Aru et al., 2012; De Graaf et al., 2012; Miller, 2007). They might be of interest to the science of consciousness broadly construed, but they aren't the central quarry of those working within the NCC framework.

Although NCCs are defined in terms of minimal sufficiency, this is a point of tension between the ideology of the NCC framework and its practice, for the methods used to identify NCCs don't allow one to draw inferences about minimal sufficiency (Bayne, 2007; Fink, 2016; Hohwy, 2009; Searle, 2000). In searching for content-NCCs, theorists contrast a situation in which a participant has one type of experience (say, a face) with a situation in which they have another type of experience (say, a house). The neural state (e.g., activity in the fusiform "face area", FFA) that is implicated in the face experience (but not the house experience) is then described as 'the NCC for faces'. However, this paradigm clearly provides

[1]Although most theorists distinguish between these different types of NCCs, there is disagreement/confusion about how they are related to each other. For example, Koch et al. state that the NCCs for generic consciousness (what they call the 'full NCCs') can be identified with "the union of the sets of content-specific NCC for all possible contents of experience" (Koch et al., 2016, p. 308). That suggestion is mistaken, for if there is an *a priori* connection between generic-NCCs and content-specific NCCs it is likely to involve the intersection relation rather than the union relation. However, in our view the NCC framework is best developed without appealing to any particular view of the relationship between different types of NCCs, for different accounts of the structure of consciousness will entail different relations here.

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60

©The author(s). https://philosophymindscience.org ISSN: 2699-0369

no reason to think that FFA activity is fully sufficient for face experiences, for it doesn't allow one to screen off neural activity that is necessary both for experiences of houses and for experiences of faces. Further, there are good reasons to doubt whether FFA activity could suffice for an experience of faces. After all, one wouldn't expect to be able to generate face experiences by placing a section of FFA in a petri dish and running some current through it. All one can conclude from this kind of contrastive experiment is that FFA suffices for face-experiences in an awake conscious participant with an otherwise complete and intact brain.

In response to this point, some theorists distinguish between a conscious state's total NCC (which guarantees its existence) and its core NCC, which is "the part of the total NCC that distinguishes one conscious state from another – the rest of the total NCC being considered as the enabling conditions for the conscious experience" (Block, 2005, p. 47; also Chalmers, 2000). With this distinction in hand (it is claimed) we can then say that the contrastive method just outlined allows one to draw inferences only about the core component of a content-NCC.

Although the NCC framework clearly requires something like the distinction between total NCCs and core NCCs, it is less clear what exactly the distinction amounts to or how it might be drawn. In distinguishing between core and total NCCs Block and Chalmers draw on Shoemaker's (1981) distinction between the core and the total realizer of a functional state. However, neither Block nor Chalmers espouse functionalism, and it's unlikely that they intend for the distinction between core and total NCCs to be understood in functional terms. Block suggests that the non-core component of an NCC can be viewed as an enabling condition for the conscious experience, but that suggestion can be understood in two ways. On one view an enabling condition is merely a causal factor in the generation of the core component. However, that doesn't seem to be what Block has in mind, for that view would imply that activating the core NCC in a petri dish would generate an experience – a state of affairs that he clearly rules out. Another possibility is that the non-core component is a background condition: its role isn't to activate the core component, but to collude with the core factor in generating a conscious state in the way in which the striking of a match colludes with oxygen to generate fire. But if this interpretation is right (and it is arguably what Block and Chalmers have in mind), then the choice of labels is a somewhat unhappy one, for the contribution of the non-core NCC is no more marginal or peripheral than the activity of the core NCC. The distinction between core NCCs and non-core NCCS might be important from an *explanatory* point of view (with core NCCs being central to some explanatory projects and non-core NCCs central to others), but it is not clear what metaphysical ice it would cut.

Arguably the fundamental distinction that Block and Chalmers are driving at here is the distinction between neural activity that is specific to a particular type of content and that which is non-specific – i.e., which occurs across a range of conscious contents. In other words, core NCCs are essentially *differentiating* neural factors, whereas non-core NCCs are non-differentiating factors. The

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60


©The author(s). https://philosophymindscience.org ISSN: 2699-0369

differentiating/non-differentiating distinction is certainly useful – not least because it maps on to the practice of consciousness science – but arguably it undermines the official definition of content-specific NCCs. Recall that the NCC for *C* is defined as the minimally sufficient basis for *C*, but we've seen that the focus of consciousness science concerns a component of that NCC that *isn't* minimally sufficient for it. Given that the science of consciousness actually focuses on core NCCs, one has to ask whether the science of consciousness is well-served by a framework that has as its ostensive focus the search for minimally sufficient neural states.

Given the problems that we have identified, why describe the search for the mechanisms/neural substrate of consciousness in terms of the 'correlates of consciousness'? The terminology of correlates is largely motivated by a desire to remain neutral on contested questions about the fundamental nature of consciousness (Chalmers, 2000; Crick, 1996). Some theorists take conscious states to be *identical* to brain states, others regard conscious states as *constituted* or *realized* by brain states, and still others hold that conscious states are merely *correlated* with brain states. By not taking sides in the disputes about the ultimate metaphysical relationship between brain states and conscious states, the NCC framework is ideally suited for the science of consciousness, for it seems to be perfectly anodyne and contests no contested issues.

Metaphysical neutrality is indeed desirable, but it is not the only kind of neutrality we should be looking to preserve. Another kind of neutrality worth preserving is *structural* neutrality: insofar as it is possible, the framework of consciousness science should avoid taking sides in debates about the structure of consciousness. In this regard the NCC framework is problematic, for it presupposes a particular conception of the structure of consciousness, as we will now show.

3 The Lite-Brite model and the neural correlates of consciousness

A Lite-Brite is a popular children's toy consisting of a light box with translucent coloured plastic pegs (see Fig.1 for a representation). These pegs – which come in nine vibrant colours – are fitted through slots in black paper in the light box. When illuminated, a picture emerges that is a function of the colours and locations of the pegs. What we'll call the *Lite-Brite model of consciousness* treats the structure of consciousness as similar to that of the eponymous toy.

The Lite-Brite model has three key commitments.[2] The first is the *Autonomy Assumption:* whether or not a neural event functions as the basis of a particular

[2]There is a close parallel between the Lite-Brite model and what Searle refers to as the building block model of consciousness (Searle, 2000; see also Bayne, 2010). There is also a close parallel between our criticisms of the Lite-Brite model and the criticisms of the NCC framework developed by Noë and Thompson (2004) and Neisser (2012).

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60

©The author(s). https://philosophymindscience.org ISSN: 2699-0369

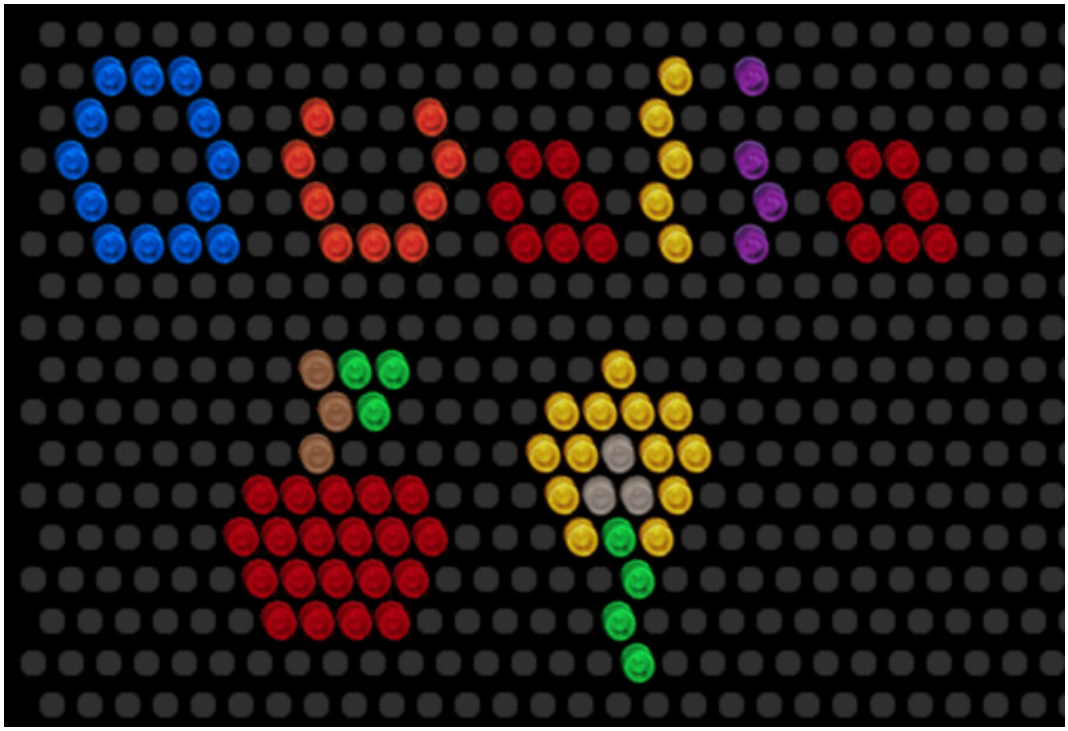

Figure 1: A representation of a virtual Lite-Brite

conscious experience does not depend on the nature or presence of other experiences had by the subject at the same time.[3] This assumption has a parallel in the structure of the Lite-Brite, for the colour that is instantiated at any particular location is independent of the colours that are instantiated at other locations (or, indeed, of whether any other colours are instantiated in the Lite-Brite at all).

The second key commitment of the Lite-Brite model is the *Uniqueness Assumption.* Uniqueness says that the presence/absence of a certain type of conscious state is determined by the activity of a single neural element. Given domain-general enabling conditions, an experience of (say) a face will depend on activity in a single face NCC. Again, there is a parallel with the structure of a Lite-Brite, for the colour at any one location in the figure is determined by a single element in the light box, and to change its color there's only one place to intervene.

The third commitment is the *Uniformity Assumption:* global changes to a subject's conscious state have uniform effects on their experiences, and, crucially, do not change the relationship between a particular type of experience and its neural basis. For example, the assumption would be that a transition from wakefulness to dreaming would uniformly affect all parts of consciousness, and not differentially alter the relation between the neural substrates and different dimensions of conscious experience. Here too there is a parallel with the structure of a Lite-Brite, for

[3]There is a parallel here with the pure insertion assumption in cognitive neuroscience (Friston et al., 1996; Sternberg, 1969).

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60

©The author(s). https://philosophymindscience.org ISSN: 2699-0369

there is a robust distinction between the specific changes made by the pegs and the general changes made by the bulb. Making global changes to the Lite Brite (by, for example, changing the wattage of the bulb), will have uniform effects on each of its components.

The conception of consciousness that emerges from these three assumptions is beguiling in its simplicity. If Autonomy is correct, then one can look for the neural basis of a particular type of experience *C* without worrying about what else one's participant might be experiencing. All that matters, it seems, is that they are having a *C*-type experience. If Uniqueness is correct, then one need find only a single neural state in order to account for *C*. Having found such a state, one's job is done. And if Uniformity is correct, then one need not worry about controlling for differences in global states. If one has found the neural basis of *C* as it occurs in (say) the state of alert wakefulness, then one can be confident that one has also found the neural basis of *C* as it occurs across every other global state.

It should be evident that the Lite-Brite model informs the NCC framework, for all three of these assumptions are embraced by the NCC framework. Although that commitment is rarely explicit, it can be seen in the language of the NCC framework and in the kinds of experiments that people conduct – or fail to conduct.

Consider first Autonomy. Theorists rarely (if ever) consider the possibility that a neural state might be sufficient for *C* in some contexts but not others. In the same way that the colour produced by a particular peg in the Lite-Brite is independent of its context, so too it is typically assumed that the experiential state produced by an NCC will be independent of its context. In other words, the NCC framework routinely ignores the possibility of interaction effects between the various components that jointly constitute the neural basis of consciousness. Instead, the NCC for an individual's total conscious state (their overall subjective perspective) is assumed to be a simple conjunction of the NCCs of each of the experiences that make up that total state.

The NCC framework is also committed to Uniqueness. Consider the following representative quotation:

> The content-specific NCC are the neurons (or, more generally, neuronal mechanisms), the activity of which determines a particular phenomenal distinction within an experience. For example, the NCC for experiencing the specific content of a face are the neurons that fire, on a trial-by-trial manner, whenever a person observes, imagines or dreams a face, and are silent in other circumstances. (Koch et al., 2016, p. 308)

Because NCCs are understood to be merely sufficient for the corresponding conscious state (and not also necessary), the NCC framework allows that different kinds of neural states could qualify as NCCs for a particular kind of conscious state in creatures of different kinds. That being said, it is widely assumed that for each state of consciousness there will be a single NCC in neurotypical humans.

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60


©The author(s). https://philosophymindscience.org ISSN: 2699-0369

This assumption of Uniqueness is reflected in the standard phrasing of *the* NCC for an experience, rather than *an* NCC or *one of many* NCCs.

Finally, the NCC framework is committed to Uniformity. Theorists who operate within the NCC framework rarely (if ever) consider the possibility that there might be one set of neural correlates for (say) *C*-as-it-occurs-in-wakefulness and another set of neural correlates for *C*-as-it-occurs-in-dreaming. Note that this is an analogue of Autonomy in the case of global states rather than contents. However, it is important for the NCC framework to preserve this distinction alongside Uniformity, lest the core/total distinction break down.

We have suggested, then, that the NCC framework is tightly wedded to the Lite-Brite model, such that the failure of the Lite-Brite model would raise real questions about the suitability of the NCC framework. Of course, that implication would be unproblematic if we had good reason to embrace the Lite-Brite model. Do we?

We don't believe so. For one, there are numerous examples of complex scientific systems which would not fit the corresponding Lite-Brite model in their domains. Sciences like genetics advanced precisely because they gave up on a general commitment to a simple 1-to-1 mapping between genes and phenotypic traits. Cognitive neuroscience has increasingly moved away from a simple view of cognitive ontology (Anderson, 2014), and we should expect the underpinnings of consciousness to be no less complex than cognitive domains. We will return to this point shortly.

Furthermore, few of the leading theories of consciousness are committed to – or even support – the Lite-Brite model. For example, one of the leading accounts of phenomenal character equates it with representational content of a certain kind (e.g., Dretske, 1997; Tye, 1995). Although one could equate the relevant kind of representational content with the content of a specific neural state (the 'vehicle' of that experience), there is nothing in the representationalist treatment of phenomenal character which mandates that position. Arguably, the only influential theory of consciousness that is committed to the Lite-Brite model is the identity theory, famously championed by J.J.C. Smart in the 1950s and occasionally still defended by philosophers (e.g., Hill, 1991; McLaughlin, 2007; Polger, 2006).

The advocate of the Lite-Brite model might be tempted to brush aside these points and insist that we already have robust evidence in favour of the Lite-Brite approach. Don't lesion, TMS and neuroimaging studies show that the FFA is the NCC for visual experiences of faces? That V4 is the NCC for colour? That MT/V5 is the NCC for motion perception? We don't deny that certain areas of the brain are more centrally involve in some types of conscious states than others. But acknowledging that point falls a long way short of endorsing the Lite-Brite model as a general account of consciousness. Furthermore, it is perhaps not surprising that the successes of the science of consciousness (such as they are) have a 'Lite-Brite' flavour, for few theorists have explored the empirical merits of the interactionist alternative. The fundamental point here is not that the Lite-Brite model is false,

©The author(s). https://philosophymindscience.org ISSN: 2699-0369

but that the NCC framework is systematically blind to the possibility of the interactionist picture.

In sum, there is ample reason to develop a framework for the science of consciousness that is not committed to the Lite-Brite model in the way in which the NCC framework is. That is the task of the next section.

4 Explanation and difference-making relations

As we have noted, the search for the neural basis of consciousness is often identified with the search for the neural mechanisms of consciousness. Yet with a few exceptions (Hohwy & Frith, 2004; Irvine, 2013; Neisser, 2012), there has not been much attempt to say what is distinctive and useful about mechanistic explanation for the science of consciousness. We suggest that the attractive component of mechanistic explanations – and the one that consciousness science ought to adopt – is the idea that explanation is done by citing not correlations but *difference-making relations* (DMRs).

At the core of a DMR is the idea that manipulating one aspect of a system enables one to manipulate other aspects of it. It is the identification of these manipulability relations, and not the identification of correlations, which enables one to understand, predict and intervene on the operations of mechanisms (Craver, 2007). Evidence about difference-making relations can be the result of direct (e.g. TMS), indirect (e.g. having different treatment conditions), or serendipitous (e.g. lesion studies) manipulations.[4]

Moving towards a difference-making account obviates the need to posit 1-to-1 relations between activities of parts and *explananda.* This has been explicitly noted as an advantage of difference-making accounts in mature sciences (Klein, 2017; Sterelny & Kitcher, 1988). Discussing the practice of labelling genes *for* phenotypic traits, Sterelny and Kitcher remark:

> Consider the vast number of loci in Drosophila melanogaster which are labeled for eye-color traits – white, eosin, vermilion, raspberry, and so forth. Nobody who subscribes to this practice of labeling believes that a pair of appropriately chosen stretches of DNA, cultured in splendid isolation, would produce a detached eye of the pertinent color. Rather, the intent is to indicate the effect that certain changes at a locus would make against the background of the rest of the genome. (Sterelny & Kitcher, 1988, p. 348)

[4]The mechanistic account is thus facilitated by the difference-making associated with manipulations and interventions. Note that it is not necessary to appeal to mechanisms in order to formulate explanations in terms of difference-making via interventions. That is, difference-making can be described without appeal to mechanistic notions of parts and processes underlying phenomena of interest.

©The author(s). https://philosophymindscience.org ISSN: 2699-0369

DMRs between different genes and phenotypic traits provide an important initial source of data for genetic theory, but the mapping from gene to trait is rarely one-to-one. Many traits are affected by multiple genes, and many genes simultaneously affect multiple traits.

DMRs are not typically found by looking for simple correlations between phenomena of interest and underlying mechanisms. Instead, one first catalogues the myriad and complicated DMRs that obtain, and then builds a mechanistic theory that accounts for them. Consider, for example, primary ciliary dyskinesia. This is a genetic disorder associated with upper respiratory infections, infertility, and a 50% chance of having the heart on the right side of the body. This surprising cluster of defects is due to a single mutation that affects the development of cilia, and so sheds light on the mechanisms of symmetry-breaking in the developing foetus and the development of the tails of sperm. The fact that a single gene affects a number of otherwise distinct variables also suggests unexpected mechanisms common to each.

Two further features of DMRs are worth noting. First, DMRs make a difference relative to a contrast class. Different contrast classes will result in different explanatory factors. To explain why someone has a fever *rather than not* we might best cite the presence of the influenza virus; why they have a *severe fever* rather than a *mild one* by reference to features of the individual's immune system; why they have a *remittent fever* rather than a *constant fever* by reference to the hypothalamic regulatory system, and so on.

Second, DMRs can vary in strength across several axes. Some difference-making relations are basically switch-like: they toggle their target between two possible states. Other DMRs are more specific*:* they connect numerous states of the difference-maker with numerous states of the target in a roughly one-to-one fashion, like the tuning dial on a radio (Woodward, 2010). Similarly, some difference-makers give a systematic handle on their targets, as they allow for well-defined responses to change (Klein, 2017). Contrast this with relationships such as 'being minimally sufficient for', which are basically binary: either the relationship obtains or it doesn't, and there's little more to say.

5 Difference-makers for consciousness

Given the success of the search for DMRs in other fields, we suggest a similar approach to consciousness. We'll call this the 'DMC framework', for it places the search for the *difference-makers of consciousness* (DMCs) at the centre of consciousness science.

At the heart of the DMC framework is the search for control variables whose settings have systematic effects on consciousness. If we change these settings, then we should expect certain effects on consciousness. More formally, let E be a variable which corresponds to some particular conscious state of interest, which can take values $\{E_1, E_2, \ldots, E_n\}$. Intuitively, if E designates a particular determinable

©The author(s). https://philosophymindscience.org ISSN: 2699-0369

experience, the different possibilities correspond to different determinates of that determinable.[5] So for example, if E corresponds to a visual experience of color, the different options $\{E_1, E_2, \dots, E_n\}$ might correspond to different hues or levels of brightness or saturations of that color. Similarly, let R correspond to the pattern of neural activity in some region of interest, which can take values $\{R_1, R_2, \dots, R_n\}$. ("Pattern of neural activity" and "region of interest" are placeholders here; the framework is compatible with most things you might want to plug in on the brain side).

Then we can formalize the picture as follows: R is a DMC with respect to E just in case there is an intervention on the value of R which changes the value of E. Thus, intervening on V4 (while keeping everything else relevant fixed) in a certain way might (e.g.) change the character of an aspect of one's visual experience from one hue to another.

The fundamental difference between the NCC and DMC frameworks thus stems from the difference between a correlate and a difference-maker. Correlates are two-place relations (R is a correlate of E), whereas difference-makers are four-place relations: R_1, *rather than* $\{R_2, \dots, R_n\}$ makes it that E_1, *rather than* $\{E_2, \dots, E_n\}$. The added flexibility of the DMC framework comes in part because of the additional argument places and the possibilities they create.

One place where the difference between the two frameworks is most stark is when properties of a single perceptual experience might depend on combinations of underlying states. This is often the case with neural population codes. So suppose, for example, that faces are represented neurally as a weighted combination of eigenfaces (Tsao & Livingstone, 2008). This sort of population coding is one in which the whole population is necessary for the representation, but in which different gestalt aspects of the represented face are affected by different weightings. So for example, tweaking along one dimension of the representation might make the face look younger or older, tweaking along another dimension might make the face look more stereotypically masculine or feminine, and so on.

On the NCC approach we might note that activity R_1 correlates with face experience E_1. We might also note that different patterns might correlate with different experienced faces. The DMC approach, by contrast, allows us to capture the combinatorial aspect of the population coding. We might show, for example, that pattern R_1, *rather than* R_2, means that the face looks young *rather than* mature; R_1, *rather than* R_3 makes the face look heavyset rather than slender; R_1, *rather than* R_4 makes the face look friendly rather than suspicious, and so on. In short, the *same* population can support various difference-makers for different aspects of an experience. Similar combinatorial principles can also apply across brain re-

[5] Woodward discusses the determinable-determinate relation as holding between variables and their values (Woodward, 2003, p. 39). In the cases we discuss, that relationship holds because of a determinable-determinate relationship between what is designated by the terms and values. Following Funkhouser (2006) we will also assume that a single determinable might have more than one dimension of determination.

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60


©The author(s). https://philosophymindscience.org ISSN: 2699-0369

gions, giving straightforward violations of Uniqueness. So for example we might find that activity in R makes the difference between seeing a face *rather than a house*, while activity in S makes the difference between seeing a face *rather than* an arm.[6]

The extra argument places afforded by a move to DMCs thus allow for a theory which respects combinatorial relationships between neural processes and conscious experiences. In other realms where similar combinatorial relationships hold – such as the relationship between genotype and phenotype – the search for DMRs brings needed clarity. Given the ubiquity of combinatorial coding schemes in the brain, one ought to expect similar benefits for the science of consciousness.

A further advantage of the DMC framework is that it permits a more nuanced picture of the relationship between global states and particular conscious contents. First, any DMR holds only relative to some background conditions and over some range of possible interventions (Woodward, 2003, pp. 239–315). Hence the DMC framework is fine with blockers and context-sensitive difference-makers: there is nothing more remarkable here than the fact that tuning the knob on the radio manipulates the station between 88 and 108 MHz if the selector is flipped to 'FM' and between 540 and 1600 KHz if the selector is on 'AM'. Indeed, context-sensitive relations of this kind are central to understanding the structure of an underlying mechanism. Hence the DMC framework is compatible with cases where Uniqueness or Autonomy *do* hold.

On the other hand, the DMC is well-positioned to capture true interaction effects: that is, cases where the *conjunction* of two states can have effects that neither conjunct alone can have. So for example, suppose that a certain experience of suffocating paranoia *E* arises only when someone believes the police are at the door (content-bearing brain state *R*) and is under the influence of marijuana (global state *S*). The conjoined state *R&S* is a difference-maker for *E*, and systematic investigation will show how *E* can be varied or eliminated by changing either *R* or *S*. However, both Autonomy and Uniformity are violated. Autonomy is violated because although *R* is a content-bearing state, its relationship to *E* depends on, and mediated by, *S*. Uniformity is violated because changes to *S* will also generate widespread changes *of varying kinds* to other phenomenal states.

The NCC approach handles state-dependence of various sorts by insisting on a distinction between state and content NCCs, or between core and total NCCs, or more generally between NCCs for particular experiences and very general global enabling conditions for those particular NCCs to be in place. We have suggested that such a sharp distinction might be misplaced, and that in any case the science

[6]This notion of a combinatorial realizer should be distinguished from the simpler variety of combination present in conscious binding. If activity at N_2 generates blue experience and activity at N_3 generates motion, then these two forms of activity might be combined in a binding process to generate an experience of a moving blue square. Binding is compatible with the NCC project, whereas the type of combination we describe here is not. The combinatorial principles we describe do not pick out distinct aspects of an experience that can be added together – instead, the *same* aspect is the result of a combination of different underlying factors.

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60


©The author(s). https://philosophymindscience.org ISSN: 2699-0369

of consciousness need not and should not *assume* such a distinction. The extra argument places afforded by the DMC relationship give it the flexibility to deal with relationships in interactive, combinatorial systems – and we have every reason to believe that the neural underpinnings of consciousness constitute such a system.

6 The difference difference-makers of consciousness make

We turn now to consider a number of possible objections to the DMC framework.

First Objection: "Isn't the contrast between the NCC framework and the DMC framework relatively superficial, and really just a matter of notation? Sure, we could talk in terms of DMCs rather than NCCs – but what precisely would we gain by so doing? As far as the science of consciousness is concerned, adopting the DMC framework would surely be business as usual."

Reply: We think that this objection arises because of a tacit commitment to a Lite-Brite structure for consciousness. Although difference-making relations are always potentially plural and contrastive, in a Lite-Brite this point is academic. The colour at any point is solely determined by the peg (if any) that is fitted: contrasts, contingently, do not come apart. So *if* consciousness has a Lite-Brite structure, then the DMC framework collapses into the NCC one – the extra argument places, while present, do no interesting work. However, we have suggested that the Lite-Brite model is false and, more importantly, that consciousness science ought not *assume* that it is true. One cannot build a theory of interactionist systems by studying single contrasts. Instead, one must study a range of contrasts, including sophisticated ones, and studying more sophisticated contrasts and intermediate cases between causal factors and background conditions. That in turn should affect the types of experiments we run and the experimental conditions we consider.

There is a grain of truth to this objection insofar as adopting the DMC framework needn't demand a revolution in the *methods* of consciousness science. However, this isn't because the contrast between the NCC and DMC frameworks has no practical implications, but because the vast majority of studies that consciousness scientists conduct are best understood as attempts to identify difference-makers. For example, although the contrastive method – comparing the brain of someone in conscious state E_1 with someone in E_2 – is clearly unable to tell us what the minimally sufficient neural basis of either E_1 or E_2 is, it is able to reveal an important difference-maker for E_1-rather-than-E_2. Hence if it seems like there is little daylight between DMCs and NCCs, that is likely because much of consciousness science is *already* looking for DMCs rather than NCCs. Conversely, we suspect that any attempts to clarify the NCC approach in ways that make it consistent with actual practice will end up simply re-introducing a version of DMCs. Hence the DMC approach is not merely a notational variant of the NCC approach: rather,

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60

©The author(s). https://philosophymindscience.org ISSN: 2699-0369

the NCC framework is an inadequate attempt to capture what the DMC approach does well.

Second Objection: "Doesn't the DMC framework make consciousness science too easy? After all, difference-makers are thick on the ground. An appropriately illuminated banana is a difference-maker for experiences of yellow, is it not? If we're only looking for difference-makers then surely consciousness science is trivially easy."

Reply: There are two responses to this objection: one conservative, the other liberal. To start, note that we have so far remained neutral between difference-making considered in the contexts of *causal* relationships and in the context of *constitutive* relationships. The two relations are distinct (the former is diachronic, the latter synchronic) but the logic of difference-making is similar in each. The extension of difference-making from causal to constitutive relationships has not been without controversy; we are on the side of those who think that it is ultimately unproblematic.[7]

The conservative response is simply to restrict the science of consciousness to constitutive (and so presumably neural) DMCs. External states such as the presence of a well-lit banana don't count – the thought runs – because they are merely causal difference-makers. Miracchi's (2017) arguments in favour of 'generative' relationships for conscious realizers could be read as an instance of this strategy.

The liberal response would be to place no restrictions on the type of difference-maker, and instead appeal to differences between difference-makers to guide consciousness science. Recall that difference-making relationships vary along a number of axes: for example, they can vary in specificity, systematicity, and robustness. These are not variations in quality *per se*: However, in general the more specific, systematic, invariant, robust difference-makers give more information about underlying mechanisms. So the presence of a banana *is* a DMC for color, but it is not a particularly informative one: many variations in the banana don't affect the color experience, the same banana gives rise to different experiences given different illuminants, the banana may make no difference at all if the subject is distracted, and so on. The reason to focus on (say) V4 is that we assume that it has a tighter relationship to experienced color under the same circumstances: we could change color experience by manipulating it even if everything else were kept fixed.

On the whole, we find the liberal response more attractive. We are convinced that part of the power of the DMC framework comes from its inclusive breadth. This may include both causal (Neisser, 2015) and even cross-level (Janssen, 2019) relationships. Here, as elsewhere, it is to the advantage of the DMC framework that it begs no contested questions about the location of the physical substrates of

[7] Craver (2007) notably used a variety of constitutive intervention to defend mechanistic explanation. There has been a substantial debate about the degree to which causal and constitutive explanations differ. Some of this debate is about the metaphysics of constitutive interventions as such, but much of it is about whether these are compatible with other assumptions of the mechanistic framework. Kästner and Andersen (2018) provide an excellent overview of the state of the debate.

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60

©The author(s). https://philosophymindscience.org ISSN: 2699-0369

consciousness (c.f. Kaplan, 2012, for a similar argument). Further, consciousness science can and does benefit from the sort of careful, systematic study of stimuli done by perceptual psychophysics. So the mere fact that something is external to the head does not mean that it is not a useful DMC for building a science of consciousness.

Either way, it is important not to fall back into assuming a Lite-Brite model. The DMC framework does not insist that we look only for maximally specific (etc.) DMCs. All DMCs are potentially relevant to building a theory of conscious experience. That may include DMCs which have surprising non-specific, non-selective effects on consciousness. If, for example, the claustrum functions as a kind of master switch for consciousness (Crick & Koch, 2005), that is useful information to know. It is useful in part because it might guide future research into the effects on consciousness of *other* interventions on the claustrum, such as those due to psychedelic drugs (Stiefel et al., 2014).

Third Objection: "But now the DMC framework is in danger of making the science of consciousness too hard! Given the complexity of the brain, there is no possibility whatsoever of being able to carry out ideal/surgical interventions. If that's the holy grail of the DMC program then the science of consciousness is doomed to failure."

Reply: It is important to distinguish DMCs themselves from evidence concerning DMCs. Claims about difference-makers are claims about what would happen were we to make an *ideal intervention:* that is (roughly speaking) an intervention on a brain region *R* which has an effect on *E* only via the *R*-to-*E* pathway, while everything else is kept fixed. A host of practical problems makes it difficult to see how an ideal intervention could ever be done. Constitutive DMCs in particular require special treatment if they are to meet some plausible analogue of this condition.[8]

Yet the search for DMCs is the search for difference-makers that *would* behave in the right way *were* we able to make ideal interventions. It is not the search for ideal interventions, and does not require them. All kinds of evidence might be relevant to establishing claims about DMCs. In the happiest scientific cases, the evidential relation is simple and direct: we establish a DMR by making an ideal intervention and seeing what happens. There are straightforward experiments in e.g. physics, or medicine, or agronomy, that admit of such manipulations. In most cases, however, we are searching for evidence about DMRs that comes from making non-ideal interventions on systems which can't be held fixed.

Much of the debate is thus over how non-ideal evidence might justify claims about difference-makers. Similarly, contrastive studies in neuroimaging might provide evidence about DMRs for cognition (Klein, 2017), but that evidential relationships requires a fair number of background assumptions about the technical

[8]This is the core of the Baumgartner and Gebharter's (2016) 'fat-handedness' objection to simple constitutive notions of difference-making. See Kästner and Andersen (2018) for a review of subsequent debates on fat-handedness.

©The author(s). https://philosophymindscience.org ISSN: 2699-0369

details of fMRI, the underlying hemodynamic response, and so on. This is not unique to the study of consciousness. Economists (say) might want to know what would happen if you raise the minimum wage but kept everything else as fixed as possible. That must be extrapolated from a bunch of decidedly non-ideal policy interventions. Similarly so, we suggest, in the science of consciousness.

Fourth objection: "The science of consciousness is in the business of providing explanations of consciousness, but difference-makers are not necessarily explanatory. How, if at all, does the DMC framework hope to explain consciousness?"

Reply: The question of whether DMCs are explanatory depends in part on what one requires of an explanation of consciousness. In some sense, an appeal to a difference-maker is explanatory, for difference-makers answer 'why' questions. *Q*: "Why did this person have a C_1 experience and not a C_2 experience?" *A*: "Because they were in neural state N_1. Had they been in *N2* then (all other things being equal) they would have had C_2 experience." Of course, this answer doesn't provide an explanation for why N_1 is a difference-maker for C_1 (relative to C_2), for adverting to difference makers doesn't itself explain difference-making relationships. But in this sense the DMC framework is in much the same boat as the NCC framework, for although the NCC framework invokes correlations to explain conscious states those correlations themselves remain unexplained.

What the objector really has in mind when she asks for an explanation of consciousness is an explanation of why consciousness is characterized by certain DMRs rather than others. Why is it the case that activation of the FFA is a difference-maker for experiences of faces (and is not a difference-maker for auditory experience)? As we see it, answering this question is not the job of the DMC framework as such but is rather the job of a theory of consciousness. Deploying the DMC framework will be pivotal in the development of such a theory, but the framework is distinct from a theory of consciousness itself.[9]

How might the science of consciousness progress on the DMC picture? What is required is a two-step process. Very broadly, stage one involves cataloguing the extent and type of difference-makers for a variety of conscious phenomena. Stage two involves introducing a theory about why those difference-making relationships hold. These two stages are conceptually distinct, but are likely to be temporally intertwined in practice: theories guide our search for difference-makers, which in turn refine theories, and so on.

The first stage involves cataloguing difference-makers in as many different and useful ways as possible. The DMC approach encourages breadth in at least four different ways. First, it encourages researchers to look for multiple ways to manipulate different aspects of the same experience. Second, it encourages the search for

[9]There are intriguing remarks which suggest that generalizations themselves might be explained (see e.g. Woodward, 2003, pp. 13–14) via an interventionist process. As difference-making generalizations aren't completely invariant, one might have a picture on which the generalizations themselves can be varied in systematic ways. Whether one explains a generalization or merely parameterizes an existing one is a complex question which in part on the difficult problem of variable choice (Woodward, 2016). This issue is explored further in Klein & Barron (2020).

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60


©The author(s). https://philosophymindscience.org ISSN: 2699-0369

interventions that change a variety of conscious contents in correlated ways (as, for example, might happen with psychedelics). Third, it encourages the search for more complex relationships (such as specificity and systematicity) between neural states and conscious contents. And fourth, it encourages breadth in the search for types of difference-makers on the neural end, including difference-makers at a variety of levels of organisation.

Stage two of the process is abductive. It involves building a theory which explains the difference-making relations discovered in stage one. There is comparatively little in the difference-making literature about how difference-making relations themselves are to be explained, and this is properly a project for a different paper. The important thing is that the *shape* of this explanatory story will look very different to that of stage one of the project. This is a general feature of mechanistic explanations.

Consider again the analogy with genetics. Stage one would involve cataloguing a variety of genes for various phenotypic traits, along with their linkages and potential variants. Other evidence about broader difference-making traits might be relevant here – the effects of radiation and various chemicals on overall mutation-rates, for example, or coordinated influences in development. Stage two involves giving a broad mechanistic story in which these individual relationships can plausibly be embedded. The story about chromosomes, DNA, transcription, and so on explains *why* the myriad DMRs identified in stage one might plausibly hold.

A similar process, we suggest, can be expected for consciousness. Finding NCCs is often presented as an end in itself. DMCs, by contrast, are a useful first step in building up a theory of why those DMCs hold – a theory-building process that we should expect to culminate in a systematic scientific theory of consciousness.

7 Conclusion: How to run with the DMC framework

Consciousness has some kind of compositional structure. A total experience involves multiple phenomenal elements, many of which can occur in the context of other total experiences. (One can taste coffee whilst working in one's office or whilst lying on the beach.) The neural substrates of consciousness also possess some kind of compositional structure, for interventions on different components of the brain can affect consciousness in different ways. It is thus tempting to suppose that there must be a straightforward mapping from the phenomenal components of consciousness to the components that make up its neural substrate. The Lite-Brite model endorses that thought, holding that the relationship between consciousness and its neural substrates is characterized by Autonomy, Uniqueness and Uniformity. And, as we have seen, the NCC framework also endorses that thought insofar as it embraces the Lite-Brite model.

Klein et al. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II), 4. https://doi.org/10.33735/phimisci.2020.II.60


©The author(s). https://philosophymindscience.org ISSN: 2699-0369

In this paper we have suggested that the relationship between the structure of consciousness and that of its neural substrates is much more likely to follow the interactionist pattern that characterizes other complex systems, in which straightforward mappings between the activity of any one component (e.g., a gene) and the presence of a high-level property (e.g., a behavioural trait) are rare. Attempts to understand such systems are invariably driven not by the quest for correlations but by the search for difference-making relations. Moving from NCCs to DMCs not only allows us to better understand what consciousness scientists actually do, it also suggests new avenues for trying to understand how the activity of brains generates the wonder of experience.

Acknowledgments
Thanks to Andrew Barron for helpful discussion. Work on this paper was supported by Australian Research Council grants FT140100422 (to CK), FT 150100266 (to TB), and DP190101805 (to TB & JH). TB is also indebted to the Canadian Institute for Advanced Research (CIFAR) for its support.

References

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain.* MIT Press.

Aru, J., Bachmann, T., Singer, W., & Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neuroscience & Biobehavioral Reviews*, *36*(2), 737–746. https://doi.org/10.1016/j.neubiorev.2011.12.003

Baumgartner, M., & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science*, *67*(3), 731–756. https://doi.org/10.1093/bjps/axv003

Bayne, T. (2007). Conscious states and conscious creatures: Explanation in the scientific study of consciousness. *Philosophical Perspectives*, *21*(1), 1–22. https://doi.org/10.1111/j.1520-8583.2007.00118.x

Bayne, T. (2010). *The unity of consciousness.* Oxford University Press.

Bayne, T., Hohwy, J., & Owen, A. M. (2016). Are there levels of consciousness? *Trends in Cognitive Sciences*, *20*(6), 405–413. https://doi.org/10.1016/j.tics.2016.03.009

Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences*, *9*(2), 46–52. https://doi.org/10.1016/j.tics.2004.12.006

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., & Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *Journal of Neuroscience*, *37*(40), 9603–9613. https://doi.org/10.1523/JNEUROSCI.3218-16.2017

Chalmers, D. J. (2000). What is a neural correlate of consciousness. In T. Metzinger (Ed.), *Neural correlates of consciousness: Empirical and conceptual questions* (pp. 17–39). MIT Press.

Craver, C. F. (2007). *Explaining the brain.* Oxford University Press.

Craver, C., & Kaplan, D. M. (2011). Towards a mechanistic philosophy of neuroscience. In S. French & J. Saatsi (Eds.), *The Bloomsbury companion to the philosophy of science* (pp. 268–292). A&C Black.

Crick, F. (1996). Visual perception: Rivalry and consciousness. *Nature*, *379*(6565), 485–486. https://doi.org/10.1038/379485a0

Crick, F. C., & Koch, C. (2005). What is the function of the claustrum? *Philosophical Transactions. Biological Sciences*, *360*(1458), 1271–1279. https://doi.org/10.1098/rstb.2005.1661

Cummins, R. (2000). "How does it work?" versus "What are the laws?": Two conceptions of psychological explanation. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 117–144). MIT Press.

De Graaf, T. A., Hsieh, P.-J., & Sack, A. T. (2012). The "correlates" in neural correlates of consciousness. *Neuroscience & Biobehavioral Reviews*, *36*(1), 191–197. https://doi.org/10.1016/j.neubiorev.2011.05.012

Dretske, F. I. (1997). *Naturalizing the mind.* MIT Press.

Fink, S. B. (2016). A deeper look at the "neural correlate of consciousness". *Frontiers in Psychology*, *7*, 1044. https://doi.org/10.3389/fpsyg.2016.01044


©The author(s). https://philosophymindscience.org ISSN: 2699-0369

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *Neuroimage*, *4*(2), 97–104. https://doi.org/10.1006/nimg.1996.0033

Funkhouser, E. (2006). The determinable-determinate relation. *Noûs*, *40*(3), 548–569. https://doi.org/10.1111/j.1468-0068.2006.00623.x

Hill, C. S. (1991). *Sensations: A defense of type materialism*. Cambridge University Press.

Hohwy, J. (2009). The neural correlates of consciousness: New experimental approaches needed? *Consciousness and Cognition*, *18*(2), 428–438. https://doi.org/10.1016/j.concog.2009.02.006

Hohwy, J., & Frith, C. (2004). Can neuroscience explain consciousness? *Journal of Consciousness Studies*, *11*(7-8), 180–198.

Irvine, E. (2013). *Consciousness as a scientific concept: A philosophy of science perspective* (Vol. 5). Springer.

Janssen, A. (2019). *Explanations in neuroimaging research* [PhD thesis]. Radboud University Nijmegen.

Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology & Philosophy*, *27*(4), 545–570. https://doi.org/10.1007/s10539-012-9308-4

Kästner, L., & Andersen, L. M. (2018). Intervening into mechanisms: Prospects and challenges. *Philosophy Compass*, *13*(11), e12546. https://doi.org/10.1111/phc3.12546

Klein, C. (2017). Brain regions as difference-makers. *Philosophical Psychology*, *30*(1–2), 1–20. https://doi.org/10.1080/09515089.2016.1253053

Klein, C., & Barron, A. B. (2020). How experimental neuroscientists can fix the hard problem of consciousness. *Neuroscience of Consciousness*, *06*(1). https://doi.org/10.1093/nc/niaa009

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience*, *17*(5), 307. https://doi.org/10.1038/nrn.2016.22

Laureys, S. (2005). The neural correlate of (un) awareness: Lessons from the vegetative state. *Trends in Cognitive Sciences*, *9*(12), 556–559. https://doi.org/10.1016/j.tics.2005.10.010

McLaughlin, B. (2007). Type materialism for phenomenal consciousness. In S. Schneider & M. Velmans (Eds.), *The Blackwell companion to consciousness* (pp. 431–444). Wiley Online Library.

Metzinger, T. (2000). *Neural correlates of consciousness: Empirical and conceptual questions*. MIT Press.

Miller, S. M. (2007). On the correlation/constitution distinction problem (and other hard problems) in the scientific study of consciousness. *Acta Neuropsychiatrica*, *19*(3), 159–176. https://doi.org/10.1111/j.1601-5215.2007.00207.x

Miracchi, L. (2017). Generative explanation in cognitive science and the hard problem of consciousness. *Philosophical Perspectives*, *31*(1), 267–291. https://doi.org/doi.org/10.1111/phpe.12095

Neisser, J. (2012). Neural correlates of consciousness reconsidered. *Consciousness and Cognition*, *21*(2), 681–690. https://doi.org/10.1016/j.concog.2011.03.012

Neisser, J. (2015). *The science of subjectivity*. Springer.

Noë, A., & Thompson, E. (2004). Are there neural correlates of consciousness? *Journal of Consciousness Studies*, *11*(1), 3–28. https://www.ingentaconnect.com/content/imp/jcs/2004/00000011/00000001/1400

Odegaard, B., Knight, R. T., & Lau, H. (2017). Should a few null findings falsify prefrontal theories of conscious perception? *Journal of Neuroscience*, *37*(40), 9593–9602. https://doi.org/10.1523/JNEUROSCI.3217-16.2017

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge University Press.

Polger, T. W. (2006). *Natural minds*. MIT Press.

Searle, J. R. (2000). Consciousness. *Annual Review of Neuroscience*, *23*(1), 557–578. https://doi.org/10.1146/annurev.neuro.23.1.557

Shoemaker, S. (1981). Some varieties of functionalism. *Philosophical Topics*, *12*(1), 93–119. https://doi.org/10.5840/philtopics198112145

Sporns, O. (2011). *Networks of the brain*. MIT Press.

Sterelny, K., & Kitcher, P. (1988). The return of the gene. *The Journal of Philosophy*, *85*(7), 339–361. https://doi.org/10.2307/2026953

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315. https://doi.org/10.1016/0001-6918(69)90055-9

Stiefel, K. M., Merrifield, A., & Holcombe, A. O. (2014). The claustrum's proposed role in consciousness is supported by the effect and target localization of salvia divinorum. *Frontiers in Integrative Neuroscience*, *8*. https://doi.org/10.3389/fnint.2014.00020

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, *31*, 411–437. https://doi.org/10.1146/annurev.neuro.30.051606.094238

Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind*. MIT Press.

©The author(s). https://philosophymindscience.org ISSN: 2699-0369

Woodward, J. (2003). *Making things happen.* Oxford University Press.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*(3), 287–318. https://doi.org/10.1007/s10539-010-9200-z

Woodward, J. (2016). The problem of variable choice. *Synthese*, *193*(4), 1047–1072. https://doi.org/10.1007/s11229-015-0810-5

Open Access
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (https://creativecommons.org/licenses/by-nc/4.0/), which permits noncommercial use, distribution, and reproduction in any medium, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

©The author(s). https://philosophymindscience.org ISSN: 2699-0369