How to think about the functions of consciousness

Joshua Shepherd (ICREA & Facultat de Filosofia i Lletres, Universitat Autónoma de Barcelona)

and

Tim Bayne (School of Philosophy, History, and Indigenous Studies (SOPHIS) & Monash Centre for Consciousness and Contemplative Studies (M3CS), Monash University; 'Brain, Mind, and Consciousness' program, Canadian Institute for Advanced Research (CIFAR))

A foundational issue for the science and philosophy of consciousness concerns the function(s) of consciousness – what consciousness does for any particular aspect of psychological or neural processing. In spite of progress in consciousness science, false assumptions and a lack of clarity regarding how best to approach the functions of consciousness represent an ongoing and serious roadblock to progress. Misguided approaches to the function(s) of consciousness have the potential to mangle explanatory priorities, and divert attention, effort, and funding away from useful questions and experimental paradigms. In this paper we offer a way forward: the capacity-based approach to the function(s) of consciousness. This approach flows out of a general explanatory approach that is influential in the philosophy of science and psychology (but not consciousness studies), according to which the

mind is understood in terms of a structured collection of capacities. And capacities are explained by empirically discovered facts that identify functions (causal roles) played by empirically identified parts of a system. After elucidating this capacity-based approach to the mind, we show how consciousness fits within it. We then argue that this approach avoids problems that plague theory-based approaches to identifying the function(s) of consciousness, avoids mistakes endemic to the common strategy of looking for the function(s) of consciousness by asking what consciousness is necessary for, and re-orients explanatory priorities in a way that better focuses consciousness science, and that suggests fruitful avenues for experimentation.

Keywords: functions, capacities, consciousness, mechanism

1 Introduction

A foundational issue for the science and philosophy of consciousness concerns the function(s) of consciousness – what consciousness does for any particular aspect of psychological or neural processing. This issue is foundational for many reasons: better understanding of the function(s) of consciousness promises to inform accounts of how and why consciousness evolved (Nichols and Grantham 2000); guide attempts to identify the distribution of consciousness in animals and machines (Allen and Trestman 2024; Bayne et al. 2024); and bears on the question of whether consciousness is a natural kind (Bayne and Shea 2020; Irvine 2012). But the issue remains contentious, with little agreement regarding what the function(s) of consciousness are, how they

relate to each other, how they might be identified (Ludwig 2022; Niikawa et al. 2022), or whether consciousness has a function (Robinson et al. 2015).

Most agree that the science of consciousness shows signs of growth on these issues, and that more and better data, along with improvements in measurement technology, will enhance our understanding of the functions of consciousness. In our view, however, false assumptions and a lack of clarity regarding how best to approach the functions of consciousness represent an ongoing and serious roadblock to progress. Misguided approaches to the functions of consciousness have the potential to mangle explanatory priorities, and divert attention, effort, and funding away from useful questions and experimental paradigms. In this paper we offer a way forward: the capacity-based approach to the functions of consciousness.

In sections 2 and 3 we elucidate the general approach. Section 2 focuses on a capacity-based approach to the explanation of the mind generally, and explicates the role of functions and mechanisms in this approach. Section 3 considers how consciousness fits within this general approach, and draws some conceptual distinctions critical for zeroing in on consciousness's functional roles. Sections 4-6 highlight three key virtues of the capacity-based approach. Section 4 argues that the approach avoids problems that plague theory-based approaches to identifying the functions of consciousness. Section 5 argues that in employing a mechanistic framework, the capacity-based approach avoids mistakes endemic to the common strategy of looking for the function(s) of consciousness by asking what consciousness is necessary for. Section 6 illustrates how the capacity-based approach is able to re-orient explanatory priorities in a way that allows for the integration of findings across a disparate collection of research programs.

2 Capacities and Functions

The capacity-based approach flows out of a general explanatory approach that is influential in the philosophy of science and psychology (but not consciousness studies), according to which the mind is understood in terms of a structured collection of capacities. What psychological capacities there are, how they are exercised, and how they relate to consciousness, are substantive theoretical questions. But intuitive examples would include the capacity to see, the capacity to reason, the capacity to attend, and the capacity to remember.

A key goal of the mind sciences is to accurately chart the relations between capacities.[1] Crucial, different kinds of relations need to be recognized here:

Some capacities presuppose others. For example, the capacity for ordinary visual perception presupposes capacities associated with color vision, edge detection, motion perception, object-tracking, and so on.

Some capacities relate to others as enablers. Arguably, the capacity to use external symbols and the capacity to represent serial order enable a bootstrapping process that explains the human capacity for the development of sophisticated numerical cognition (Carey 2009).

Some capacities relate to other capacities as enhancers. Arguably, the capacity for face perception enhances the capacity for emotion attribution. Although emotion can typically be attributed in the absence of face perception, face perception renders emotion attribution quicker in some

[1] Traditional box-and-arrow diagrams beloved by 'old school' cognitive psychologists (e.g. Shallice 1972), can often be understood as attempts to chart some of the relationship between capacities (or their deployment), but the kinds of relations recognized by such approaches are often impoverished, and typically fail to do justice to the rich array of ways in which capacities (and their exercise) can be related.

circumstances, and more fine-grained in others. And attention is often thought of as an enhancer of a wide range of perceptual and cognitive capacities.

Some capacities relate to others as inhibiters. For example, speech production inhibits verbal working memory (Murray 1968).[2] Given gradations in the ways that different capacities relate to each other, it is natural to expect a gradation here, with some capacities partially inhibiting others. Sensorimotor adaptation, for example, sometimes has an indirect impact upon the capacity to guide behavior, in part because the mechanisms that drive sensorimotor updating operate independently of (while maintaining sensitivity to) the planning capacities that form intentions (McDougle et al. 2016).

Many of the most active debates in cognitive science can be understood as debates about the analysis of capacities (and, often, the relations between capacities). For example, at the heart of the cognitive penetrability debate is the question of whether an agent's perceptual capacities are independent of what they believe and want (the contents of 'central cognition'), or whether perceptual capacities can be understood only in relation to what it is that an agent wants and/or believes (Siegel 2012; Firestone and Scholl 2016; Block 2023).

On a capacity-based conception of the mind it may be initially unclear where and how consciousness fits. Is the property of being conscious itself to be identified with a capacity? This way of thinking is most plausible regarding global states of consciousness ('levels of consciousness'), such as the state of waking awareness (see McKilliam 2020 for discussion). But whatever we think of global states of consciousness, it also seems that consciousness shows up in

[2] Thanks to Dan Harris for discussion here.

different ways across a wide range of mental phenomena that seem well-described in terms of the operation of capacities that can be characterized independently from consciousness: metacognition, olfaction, motor control, inhibitory control, and so on. So it is plausible to approach the functions of consciousness in terms of consciousness's impact on psychological capacities. For example, we can ask what role consciousness plays in relation to (for example) perception, thought or agency. On this picture, consciousness modulates (or modifies, regulates, or integrates) the use of another capacity (the 'target capacity'). In some cases, it is evident that the capacity in question could be had (and used) in the absence of consciousness, and the primary question is whether (and how) consciousness enables and/or enhances the target capacity. In other cases it is controversial whether the target capacity can occur in the absence of consciousness (as is the case with perception (Peters et al. 2017; Phillips 2018). Indeed, it might even turn out that consciousness inhibits certain cognitive or perceptual capacities. For example, it's been claimed (controversially!) that consciousness impairs the quality of certain forms of decision-making (Dijksterhuis et al. 2006; but see Lassiter et al. 2009). In addition to thinking about the role played by the generic property (*being conscious*) within a capacity-based conception of the mind, we need also to consider the role played by various determinate states of consciousness, such as *being perceptually conscious* or being conscious of the colour of an attended object. We return to this idea below.

Properly identifying the mind's key capacities, and charting how they relate to each other, is grounded in the further task of explaining how capacities operate. This is where functions enter the picture. Functions are causal roles played by parts of a system, that enter into explanations of how capacities work, when they work. The difference between function and capacity, on this

conception, is not intrinsic, but has to do with how theorists target mereological levels. Both functions and capacities do things, causally, for the systems into which they are embedded. On Cummins's (1975) influential formulation, functions are causal roles that contribute to explanations of the traits or capacities displayed by a broader system into which the causal role is embedded. More precisely, some part *P* of a system *S* functions to *C* in *S* if and only if *P* is capable of *C*-ing in *S*, and an (appropriate, adequate) account of *S*'s capacity *D* appeals to *P*'s capacity *C*.[3]

So capacities are explained by empirically discovered facts that identify functions (causal roles) played by empirically identified parts of a system. These causal roles, when appropriately conjoined, enable us to lay out how it is that the capacity operates – how it is that some system

[3] Philosophers of science have articulated different types of function. Although we focus here on Cummins functions, regarding consciousness, more than one type is relevant. A Wright function, for example, factors in explanations of the presence of a trait in an organism: to say the (or a) function of *X* is *Z* is to say that *X* is present (at least in part) because it *Z*s, and that *Z* is present (at least in part) as a result of *X*. So, for example, a function of the kidneys is to filter toxins: the kidneys are present at least in part because they filter toxins, and the filtration of toxins occurs as a result of the kidneys.

Wright functions are often at issue when theorists appeal to a history of natural selection in an organism. Those concerned about the function of consciousness are no exception. It would be useful to have solid evolutionary explanations of the emergence and development of consciousness – to have an explanation of 'why the functionally characterized thing exists, in the form it does' (Godfrey-Smith 1994: 344).

In the case of consciousness, such explanations are difficult to come by. One project concerns locating the emergence of consciousness in biological time. It is hard to resist quoting Susan Blackmore here, who, in her textbook on consciousness, notes that 'there is no consensus over when consciousness evolved, let alone what it is. Proposals range from billions of years ago to only a few thousand' (2012: 228).

Another project focuses instead on consciousness's recent history – on the effects of consciousness that might explain how consciousness, or aspects of consciousness, have been recently maintained under natural selection (see Godfrey-Smith 1994).

One difficulty here is the possibility of exaptation along the evolutionary line. It is plausible to think that some exaptation has occurred in the evolution of the human brain. Some argue that exaptation is closer to the rule. Anderson, for example, argues 'that cognitive evolution proceeded in a way analogous to component reuse in software engineering, whereby existing components—originally developed to serve some specific purpose—are used for new purposes and combined to support new capacities, without disrupting their participation in existing programs' (2007: 331). If something close is correct, recent history will be as (if not more) explanatorily relevant for theories of the Wright function of consciousness. Further, given exaptation, an organism could have many traits that have no Wright function, but important Cummins functions.

enables seeing, edge detection, color discrimination, reasoning, inhibition, task set construction, attending, remembering, face perception, emotion attribution, or whatever.[4]

Now, explaining how a particular capacity is realized or implemented is not yet to explain any particular exercise of that capacity. Explaining why a capacity is exercised at a time may require additional insight into features of a capacity. And explaining why a capacity is exercised well or poorly in different circumstances may call for additional research. Ultimately, we want to understand the causal structure of a capacity, as well as its performance profile across differences of circumstance, its common triggers, and typical sources of malfunction.

This approach to explaining capacities can be connected to a mechanistic approach to psychological and neuroscientific explanation, championed by Machamer et al. (2000) and Craver (2001; 2007), and developed by Krickel (2018) and Piccinini (2020), among others. On a mechanistic approach, a chief goal psychology and neuroscience share is the development of accurate descriptions of mechanisms – descriptions of the active, spatial, and temporal structure of systems of parts that illuminate how the joint operations of the parts explain causal outputs of the whole.

Since mechanisms have complex internal organizational principles, and since mechanisms are embedded in larger systems with their own organizational principles, Craver notes that a major part of mechanistic explanation is interlevel integration against a background of mechanistic

[4] One might worry that if a capacity is constituted by its causal roles (plus embedding in a broader system, perhaps) then the relationship between the causal role and the capacity is too intimate for the former to count as explaining the latter. That worry might be legitimate if the only notion of explanation here were causal, but its force dissipates once one recognizes the existence of non-causal forms of explanation.

hierarchies.[5] Interlevel integration is difficult work.[6] In the case of consciousness, it remains very programmatic.[7]

In sum, at a high level of abstraction, a capacity-based approach is thus a general approach to explaining the mind. Of course there will be significant uncertainty attached to current theorizing about many capacities. There is uncertainty attached to any proposed mental ontology. But felicitously carving up the space of capacities is an ongoing, shared goal of the mind sciences, and is connected to the identification of the mechanisms and functional roles that explain the operations of capacities.[8]

3 Consciousness and Capacities

With a capacity-based approach to explaining the mind as background, the capacity-based approach to the functions of consciousness translates the question 'what are the functions of consciousness' into the following:

[5] The extent to which mechanistic explanations can unify psychology and neuroscience remains somewhat controversial – see Weiskopf (2011) for an argument that non-mechanistic cognitive models can provide independent explanations of cognitive capacities, and Piccinini and Craver (2011) or Povich (2015) for arguments that such models can ultimately be integrated into mechanistic explanations.

[6] The individuation of levels within a mechanistic approach raises issues that we do not have space to address. Povich and Craver (2017) suggest the following as a way to individuate levels: '*X*'s ϕ-ing is at a lower mechanistic level than *S*'s ψ-ing if and only if *X* and its ϕ-ing are component parts and activities in *S*'s ψ-ing. A component is a part whose activities contribute to the behavior of the mechanism as a whole' (186). But what makes something a relevant 'component part' is a topic of debate in the philosophy of science.

[7] One can see movement in the direction of integration in some quarters. Dehaene et al. (1998), for example, have developed the global workspace theory of consciousness by linking data relevant to the global workspace to neuro-functional models of the brain. And proponents of higher-order theories of consciousness, while insisting that the theory should be pitched at the level of psychological states, take neural data regarding the prefrontal cortex to be relevant to the success of the theory (Brown et al. 2019).

[8] Folk psychology provides a useful starting point for the individuation of capacities (i.e. for developing a 'cognitive ontology' of capacities), but it is revisable in light of scientific investigation.

The Capacity-Based Account (CBA): For any given capacity, what relations (e.g., enabling, enhancing, inhibiting, and so on) does (some aspect of) consciousness have to that capacity?

The functions of consciousness are thus causal contributions to the functional roles of various capacities. Since capacities and causal roles are diverse and plentiful, and may occur across various levels of mechanistic hierarchies, aspects of consciousness might have diverse and plentiful causal roles, and crop up across a wide range of mechanisms, at different places in an explanatory hierarchy.

The notion of an 'aspect' of consciousness requires unpacking. We will distinguish between three aspects of consciousness: generic consciousness (consciousness-as-such), global states of consciousness, and local states of consciousness.

Generic consciousness (also known as 'creature consciousness') is simply the property of being conscious – of having some kind of subjective perspective. This is the highest-level determinable of consciousness. Global states of consciousness (also known as 'levels' or 'modes' of consciousness) characterise an organism's overall conscious condition', and are individuated on cognitive, behavioral, and physiological grounds (Bayne et al. 2016; McKilliam 2020). It is unclear what global states should be recognized, but it is widely assumed that there are alterations in global state associated with various states of wakefulness, sleep and sedation, and with certain kinds of neurological conditions (such as epilepsy) (see e.g., Bayne and Hohwy 2016; Boly et al.

2013; Laureys 2005; Overgaard and Overgaard 2010). Local states of consciousness (also known as 'conscious contents') concern the specific ways in which a creature's overall conscious state is modulated at a time. Typically, sensory local states are individuated in terms of their phenomenal character (often itself picked out in terms of objects and properties), while cognitive local states are individuated in propositional attitude terms (that is, in terms of intentional content and attitude). Thus, one might be in a conscious state of hearing a violin being played (with the associated phenomenal character), whilst also thinking that it would be good if that violin were not being played.

This point is relevant for the following reason. It is not obvious that the causal role played by some aspects of consciousness for one capacity (e.g., the role of color phenomenology for color discrimination or object detection) will be at all similar to the causal role played by consciousness for another capacity (e.g., task switching, or error detection, or metacognition). The role of color phenomenology for object discrimination may differ from the role of mental effort phenomenology for task switching, or the role of olfactory phenomenology for toxin detection. To avoid glut in our theorizing about the functions of consciousness, it is important to relativize function claims not only to specific capacities, but also to specific aspects of consciousness. And the specifics of phenomenology may matter for some capacities, and not for others.

Questions about the functions of consciousness can be raised not only about each of these aspects, but also about the relations between them.

First, we can ask about the functions of generic consciousness. What difference does being conscious make to an organism? Second, we can ask about the functions of the various global states of consciousness. For example, what difference does (e.g.) being lightly sedated make to an

organism? Third, we can ask about the functions associated with various kinds of local states. What difference does having certain types of local conscious states make to an organism?

Given the important differences between them, it's plausible to suppose that global states (as a class) and local states (as a class) involve distinct capacities and mechanisms, and thus that their functions will also differ. We might also expect that there will be important differences between local states of various kinds. For example, the functional role associated with conscious pain might be very different from that associated with conscious vision or conscious thought. Indeed, we might even expect that the functional role of a local state won't be invariant across changes of global state, but may depend on the agent's global states. In other words, we should take seriously the possibility that the function of (say) visual-experience-in-the-context-of-waking-awareness is different from the function of visual-experience-in-the-context-of-dreaming-awareness. And of course, these differences will be over and above those that derive from variations in cognitive architecture.

We can see now that CBA undermines two intuitively-attractive[9] assumptions about the functions of consciousness. The first assumption is this:

Assumption of System Invariance: The function(s) of consciousness are invariant across cognitive systems of various kinds. Thus, if consciousness has the function of enabling

[9] These assumptions are closely connected to the necessity-based approach (see section 5), and appear to drive reasoning that moves from the fact that *X* can be performed unconsciously to the claim that consciousness makes no difference to *X*.

multisensory integration in (e.g.) adult neurotypical humans, then it will also have that function more generally - that is, in systems of very different kinds.

The second assumption undermined by CBA is this:

Assumption of State Invariance: The function(s) of consciousness is invariant across conscious states of various kinds. Thus, if consciousness has the function of enabling (e.g.) visual experiences to be available for the direct control of thought and action, then it has the function of enabling (e.g.) auditory experiences, olfactory experiences, bodily sensations, and conscious thoughts to be available for the direct control of thought and action.

What accounts for the intuitive appeal of these two assumptions? A natural thought here is that if consciousness were a single phenomenon, then every instance of it would also have a uniform functional profile. And if that's right (one might worry), then rejecting these assumptions of system and state invariance would require one to reject the idea that consciousness is a 'single thing', and to thus embrace some form of eliminativism about consciousness.

Although eliminativism about consciousness is a view that we think ought to be on the table (Irvine and Sprevak 2020), it is not one that we would want to be committed to. Luckily, nothing that we have said thus far does commit us to eliminativism (or even anything in that ballpark), for it is

possible to embrace realism about consciousness without embracing the two assumptions identified above.

Let's begin with the *Assumption of System Invariance*. Even if realism implies that every instance of consciousness has invariant causal powers, it doesn't follow that every instance of consciousness will have the same functional profile. After all, the impact of consciousness on a system's capacities is determined not just by the causal power of consciousness itself but by the system's cognitive architecture. Compare: the petrol contained in each of 3 jerry cans might have the same causal power, but it will have very different effects depends on the kind of vehicles into which they are poured.

What about the *Assumption of State Invariance?* To fix ideas, contrast three conscious states with each other, where *CS1* is a conscious mood (say, feeling bored), *CS2* is a visual experience (e.g. seeing a dog), and *CS3* is a conscious thought (e.g., realizing that an argument is invalid). Although the issue is controversial, we will assume there that each of these conscious states can be treated as instances of an over-arching kind. In other words, *CS1*, *CS2* and *CS3* can each be thought of as determinates of the super-determinable property 'consciousness'.[10] Now, one might think that if this is the case, then it must follow that there will be a single functional profile that is shared by *CS1*, *CS2* and *CS3*, and that the Assumption of State Invariance must be secure.

We can see that there must be something wrong with this line of argument by considering a variant of our jerry can example. Suppose that instead of petrol, our three jerry cans contain liquids

[10] Here, we reject the influential view (e.g. Carruthers) that sensory states are conscious in one sense (roughly, the 'phenomenal consciousness' sense) and thoughts are conscious in another sense (roughly, the 'access consciousness' sense.

of different types. One of these liquids is appropriate for a particular electronic liquid cooling system, but the other two will ruin it due to differences in their chemical composition and the particulate matter that they contain (Nguyen et al. 2007). Thus, although each of these liquids will share an abstract functional profile in virtue of their liquidity, in many contexts that functional commonality will be less important than the functional differences between them. Similarly, although *CS1*, *CS2* and *CS3* will share an abstract functional profile in virtue of the fact that they are all instances of consciousness, that functional commonality may in many contexts be less striking than the functional differences between them. Given that moods, visual states and thoughts are themselves very different phenomena, we should expect the functional profile of conscious moods, conscious visual states and conscious thoughts to also different markedly from each other.

The discussion thus far intersects in interesting ways with a different approach to the functions of consciousness. This approach, which is perhaps taken most clearly by some proponents of higher-order theories of consciousness (Rosenthal 2008), posits conscious and unconscious manifestations of the same state-type T (e.g., visual), and seeks to find evidence regarding the functional difference between conscious Ts and unconscious Ts. One might take CBA to resist the idea that a state-type can be factored into conscious and unconscious varietals. But in fact CBA can remain neutral on this. In general, a capacity-based approach need not commit to the idea that the best way to discover the causal profile of consciousness is by contrasting conscious and unconscious states. It is an empirical question whether some state-type can be cleanly divided into conscious and unconscious varietals. The answer may vary across state-types. And while direct comparison of conscious and unconscious state-types may be useful for one explanatory purpose,

in some cases it may be more useful to compare a system that has a conscious T and a system that lacks T altogether.

In sum, then, CBA is grounded in a more general explanatory approach to the mind, that seeks to taxonomize the space of psychological capacities and then explain their operation in terms of the structure, operation, and causal collaboration of mechanisms. Crucially, CBA not only relativizes claims about the functions of consciousness to particular types of conscious systems, but also relativises such claims to aspects of consciousness.[11]

In the remainder of this paper we examine the signal advantages of CBA, and explore its implications for the science of consciousness. Section 4 contrasts CBA with an alternative picture of how the functions of consciousness are to be identified, while Section 5 contrasts CBA with an alternative picture of the form that an account of the functions of consciousness should take. Our final section demonstrates how CBA re-orients explanatory priorities in more fruitful directions than other approaches to the functions of consciousness.

4 The Theory-Based Approach

CBA contrasts in important ways with another – and arguably more influential – approach to speculation about the functions of consciousness: the theory-based approach. At the heart of the

[11] The capacity-based approach shares with the rest of the science of consciousness the difficulties associated with detecting the presence of (any aspect of) consciousness during the course of some capacity's exercise. But the capacity-based approach can help itself to the methodological state of the art – for example, to the rigorous use of confidence-based procedures (Michel 2022).

theory-based approach is the idea that an account of the functions of consciousness should begin with a particular theory (or set of theories) of the nature of consciousness.

A recent review of higher-order theories of consciousness, due to Brown et al. (2019: Table 1), offers a nice illustration of this approach. Brown et al. consider a range of higher-order approaches to consciousness, and offer a proposal about what – given each approach – the (or a) core function of consciousness is likely to be. For example, on the Higher-order Representation of a Representation Theory, which maintains that consciousness consists in a higher-order representation that represents oneself as being in a first-order state (e.g., a visual state of seeing red), the function of consciousness is 'Whatever the function of the right kind of higher-order representations of representations turns out to be' (Table 1). On the Perceptual Reality Monitoring Theory, it is 'Formation of subjectively justified beliefs' (Table 1). On the Multi-State Hierarchical Model of Subjective Awareness, it is the tokening of 'thoughts, beliefs, memories, feelings, and attributions about the world, and about oneself as an object and a subject' (Table 1). On the Radical Plasticity Thesis, it is 'Control of behavior and learning in novel situations' (Table 1).

Clearly, in the fullness of time, we should want a theory of the nature of consciousness to converge with a theory of the functions of consciousness. That said, we have serious reservations about the utility of looking to theory-based considerations to identify the function(s) of consciousness.

First, and most obviously, there is massive and pervasive disagreement about which theory, or which family of theories, is on the right track (for a survey of consciousness scientists that underscores this point, see Francken et al. 2022). In general, when disagreement between experts regarding some phenomenon is pervasive, one should lower one's credence in any particular theory of the phenomenon.

A second problem is that most theories of consciousness are seriously under-developed in some respect or another (Seth and Bayne 2022). Some theories are under-developed in that the core constructs to which they appeal lack the kinds of precision that would be required for applying them to challenging cases. For example, the global workspace theory (GWT) holds that an agent is conscious when it has a global workspace which supports the availability of representational content to a wide range of consuming systems. However, without a detailed account of what kinds of 'global workspaces' are consciousness-supporting we aren't in a position to apply GWT to infants, non-human animals or AI systems. Theories are often also under-developed insofar as they attempt to account only for some aspects of consciousness but not others. For example, higher-order theories offer an account of what distinguishes conscious mental states from unconscious ones, but they offer no account of what distinguishes conscious states from each other.

Third, even if we had a comprehensive theory of consciousness about which there was general consensus, we might still be uncertain about the function(s) of consciousness, for some theories of consciousness provide little to no guidance in this matter. Consider, for example, primitivist theories of consciousness, such as Russellian panpsychism. On this view, consciousness is an intrinsic property of physical entities, all the way down to the microphysical level. Since consciousness is so widespread on this view, the view can say very little about what the functions of consciousness may be. Certainly, it predicts no differences between the consciousness of a snail and that of a bonobo, although attention to the different capacities of these animals may uncover important differences. Or consider information integration theory, on which consciousness is most fundamentally (a certain amount of) integrated information. Thus stated, the theory leaves it open what consciousness may do for a given system. If the causal roles of integrated information vary

across differences in capacity, system architecture, and so on, then the theory needs to be conjoined with a huge set of empirical observations before the functional roles of consciousness will come into view.

That said, it looks as though certain (families of) theories of consciousness will offer some useful constraints on what the function(s) of consciousness might be. For example, if some version of the GWT is true then the functions of consciousness are surely related to integration and flexibility of cognitive and behavioural control. Similarly, if the local re-entry theory of consciousness is correct (Lamme 2006), then the function of visual consciousness almost certainly has something to do with visual feature binding. Thus, progress in understanding the nature of consciousness may well inform our understanding of the function(s) of consciousness. But the central point that we would emphasize here is that we shouldn't look first-and-foremost to theories of consciousness in order to identify the function(s) of consciousness. A better way forward is to begin by identifying the mind's capacities, and to ask what kind of role consciousness (and its various determinates) plays in the kinds of relations that those capacities bear to each other.

5 The Necessity Approach

In addition to contrasting with a dominant approach to identifying the function(s) of consciousness, CBA also contrasts with a dominant conception of the form that an account of the function(s) of consciousness ought to take. On this view, if any capacity is associated with consciousness, then it must be the case that consciousness is *necessary* for that capacity. On this view, a single instance in which an agent was able to execute that capacity in the absence of

consciousness would suffice to show that that capacity was not among the functions of consciousness. This view - or at least something very much like it - appears to motivate claims that X is not one of the functions of consciousness because it is possible to find instances in which X occurs but consciousness is not present. For example of such data, see work on visual feature binding (Keizer et al. 2015); multisensory integration (Mudrik et al. 2014); attention (Kentridge et al. 1999); change detection (Silverman and Mack 2006); deliberation (Dijksterhuis et al. 2006); conflict adaptation (Desender et al. 2012); (forms of) attentional control (Webb et al. 2016); and the processing of race or gender information during face perception (Amihai et al. 2011). Discovering that (e.g.) feature binding can occur in the absence of consciousness would certainly be a valuable result, but it wouldn't show that feature-binding is not a function of consciousness.[12]

We see three central problems with this view.[13] First, it employs an impoverished conception of the kinds of relations that are needed to do justice to the function(s) of consciousness. Although consciousness might indeed be necessary for certain capacities, many of the functions of

[12] The usual way of framing the relevant question is this: Can *X* be done unconsciously? This question could use greater clarity regarding the aspects of consciousness relevant to whatever capacity is in view. For, as stated, the question is ambiguous between [a] can *X* be done by an unconscious creature or human? And [b] can *X* be done by a creature that is not conscious/aware of the fact that they are *X*-ing? And [c] can *X* be done by a creature that is not conscious of the stimulus that is involved in *X*-ing (e.g., road conditions)?
Thinking of the functions of consciousness in terms of necessity also influences more abstract theorizing about the functions of consciousness. Consider Niikawa et al.'s (2020) recently proposed framework for studying the functions of consciousness. It is three-dimensional, requiring theorists to clarify [a] the target (or type) of consciousness at issue, [b] the explanatory relationship between consciousness and some function (consciousness-as-basis or consciousness-as-contributor to some function), and [c] whether the relationships between consciousness and some function are relationships of necessity, or of sufficiency. This third dimension, in particular, indicates that a chief goal of consciousness science is to determine what functions cannot be performed in the absence of consciousness.

[13] It is interesting to ask why this approach is so prominent. One speculative thought is that this approach is motivated by a desire to discover what consciousness is – the causal essence of consciousness – rather than what consciousness does. But, as we have noted, consciousness may not have a causal essence.

consciousness might be best understood not in terms of what consciousness is required for but in terms of what consciousness facilitates, inhibits, and so on.

Second, the focus on necessity obscures the fact that consciousness need not be involved in every exercise of capacity *C* to be involved in: (a) typical exercises of *C*; (b) the exercise of *C* in some special set of circumstances; (c) the exercise of *C* in 'non-standard' agents (e.g., infants; non-human animals, AI systems). The possibility of unconscious metacognition (Charles et al. 2013), perception (Shepherd and Mylopoulos 2021) or working memory (Persuh and Rue 2019) says nothing about whether consciousness might sometimes--indeed, often--be implicated in the operation of these capacities.

Third, the functional role centrally associated with some capacity might fail to qualify as one of the functions of consciousness *even if* consciousness is (in general) necessary for the exercise of that capacity. After all, something can be necessary for the exercise of a capacity while having low explanatory relevance for any account of how that capacity works. Consider an analogy: a working starter could be necessary for a car to reach 180 mph, even though the most relevant explanatory parts of the car are not the ignition system, but the engine and aerodynamics.

When theorizing about functions and capacities generally, it seems better to de-emphasize the question of necessity in favor of a focus on the relative causal contributions of multiple factors. Given functional redundancy in the nervous system, necessity may in fact be hard to come by. Additionally, some capacities will be driven by mechanisms that work probabilistically. Some capacities may be driven by a confluence of multiple factors, and effect sizes may need to be consulted for information regarding which factors are relatively important, in different

circumstances.[14] The capacity-based approach encourages a move away from the necessity approach, which is sub-optimal on independent grounds, in favor of a focus on the relative causal contributions of multiple factors, including aspects of consciousness, against the background of how various systems (in-)efficiently perform their key functions.

6 CBA and The Re-orientation of Explanatory Priorities

We have claimed that the capacity-based approach can re-orient explanatory priorities in more fruitful directions than other approaches to functions of consciousness. In this section we offer two examples of this. Consider, first, the widely held view that a key function of consciousness has to do with volition, where volition roughly denotes the initiation and control of action (see Shallice 1972; Shepherd 1994; Tye 1996). In an influential discussion, Seth notes that 'The notion that the function of consciousness is to initiate and control voluntary action has enormous appeal: We consciously think about doing *X* and then we do *X*' (2009: 290). Whether the initiation and control of action are functions of consciousness, however, has long been mired in controversy. In part this is due to experimental paradigms that sought to show that consciousness is not necessary for the initiation or control of action (Libet 1985; Wegner 2004). As we have argued, this is a mistake.

[14] In developing this point further, we might fruitfully draw on the difference-making framework recently articulated by Klein et al. (2020). As they develop it, the science of consciousness should attempt to identify difference-making relations – relations that explain why manipulation of 'one aspect of a system enables one to manipulate other aspects of it' (9). And evidence for difference-making relations can be found in many ways: they 'can be the result of direct (e.g. TMS), indirect (e.g. having different treatment conditions), or serendipitous (e.g. lesion studies) manipulations' (Klein et al. 2020: 9).
Although Klein et al. (2020) discuss how difference-making interventions and evidence might go in thinking about what consciousness is (in effect, replacing the search for NCCs), they do not address how difference-making interventions might illuminate the function(s) of consciousness.

Aspects of consciousness may be important for aspects of action control even if they are not necessary (Shepherd 2016) – a point made by many who have critically engaged with experimental paradigms pioneered by Libet and Wegner (see Brass et al. 2019).[15] But in our view additional conceptual problems plague the kinds of questions guiding inquiry regarding consciousness and volition.

First, 'volition' or 'the initiation and control of voluntary action' are messy categories that [a] combine the exercise of a wide range of capacities (e.g., perception, executive control, planning, memory, motor control) [b] involve different capacities in different circumstances, and for different kinds of actions. As such, the initiation and control of voluntary action makes a poor target for theorizing about the functions of consciousness. Recent work has been responsive to this, and highlights the sophisticated cognitive architecture that underlies action control (for example, Christensen et al. 2016; Triggiani et al. 2023). Such work underscores the points made above regarding the assumptions of system invariance and state invariance. With sophisticated capacities, we cannot expect the truth about the functions of consciousness to be simple, or invariant across differences of circumstance or capacity. Thus, if we wish to link aspects of consciousness to the exercise of capacities, we have to be clear about what capacities are in view, and how they are related to each other.

Second, much of the discussion surrounding consciousness and volition focuses on whether consciousness of any sort is involved in volition. But, as many have noted, the phenomenology

[15] A referee notes, to our mind rightly, that many of the criticisms of the Libet paradigm are friendly to our CBA to the functions of consciousness, insofar as these criticisms seek to find the *differences* that consciousness actually makes to the execution of action, even if these differences are not best found at the moment of action initiation.

associated with the initiation and control of action is complex (Bayne 2008; Pacherie 2008; Shepherd 2015). If we wish to link aspects of consciousness to the exercise of capacities, we have to be clear about which aspects (e.g., visual, multisensory, or aspects to do with intention, planning, effort, etc.) of consciousness are in view.

Our discussion of the relationship between volition and the functions of consciousness has been relatively abstract; we show now that the account developed here also has concrete applications. Our focus will be on an interesting series of experiments due to Stein and Peelen (2021), which they claim can 'distinguish between functions that do and do not require consciousness' (612).

In the critical experiment, participants were required to perform three tasks: a face localization task (whether the face was presented on the left or right of a screen), a face detection task (reporting seen or unseen using the perceptual awareness scale in face-present and face-absent trials), and a discrimination task (reporting whether the face was upright or inverted). Stimuli were presented at different temporal lengths, from 8 to 33 ms, and were then backward masked.

Stein and Peelen transformed responses on all three tasks to the same sensitivity measure (d'), to facilitate comparison, and found that participants were better at localizing and detecting faces when the face was upright, across a few presentation times. The key result regarding consciousness is that at 8 ms of stimulus presentation, with a blank of 8 ms afterwards before the mask was shown, discrimination of upright vs inverted faces was at chance, but there was still an effect of face direction on localization and detection (participants were more accurate when the face was upright). One takeaway of this result, which Stein and Peelen highlight, is that 'the face inversion effect in visual detection is (at least partly) mediated by unconscious processes' (617). A

plausible inference one might make, on the basis of this effect, is that consciousness is not necessary for processing aspects of faces.

Whether this inference is interesting, or exciting, depends upon one's explanatory priorities. For reasons we gave above, we think the non-necessity of consciousness for processing aspects of faces is not the most interesting feature of this study. Follow-ups would do better not to focus upon the necessity question, but on the difference that consciousness might make to the capacities implicated in this experiment, such as those involved in face detection and face localization.

Consider, for example, that at 17 and 25 ms of stimulus presentation, one finds a discrimination effect, such that some influence of consciousness cannot be ruled out. It is possible to see this data point as interesting primarily as a temporal parameter for the examination of unconscious processing – we have to look to stimuli presentations of shorter duration. But a capacity-based approach sees these data points as interesting for other reasons as well. Consider, for example, the hypothesis that the capacities for face detection and localization involve collaboration between unconscious and conscious processing. Exploration of this hypothesis might proceed by way of experiments that attempt to manipulate the informational parameters (the features of faces) presented to participants. Is it possible to improve unconscious processing? Is it possible to speed up the temporal window for the discrimination effect?

Interestingly, Stein and Peelen found that while increasing awareness correlates with increased accuracy at detection and localization, the face inversion effect upon these capacities is largest for middle-range times. That is, face inversion has a small effect at 8 ms, and again at 33 ms – as though, with longer presentation times, the face inversion effect begins to wash out. Experiments

that ask why this is the case might help us better explore the collaboration hypothesis floated in the above paragraph. Is it possible, for example, to manipulate the relative size (or temporal parameters of) the face inversion effect by changing the informational value of aspects of the faces presented to participants?

The suggestions just made are exploratory. This is because not enough is known about the structure of unconscious processing of faces, nor about how unconscious and conscious processing relate to each other with respect to the capacities for face detection and localization. The broader and more important point is that approaching this (very interesting) experimental paradigm from the perspective of CBA, then, motivates experiments that might, otherwise, appear less essential. This is because CBA suggests there is no shortcut to understanding the functions of consciousness: we have to experimentally explore the parameters of the relevant capacities, with close attention to the presence or absence of consciousness in different aspects of capacity-exercise. More specifically, CBA makes salient the idea that we do not fully understand the influence of consciousness on the operation of various capacities until we pay greater attention to [a] the way those capacities operate, as well as to [b] how differences in the presence and character of consciousness correlate with differences in the operation of the capacities. Greater attention to [a] and [b] suggests interesting experimental questions, which promise to bring important clues to light. But to pursue these experimental questions, it may be necessary to place the functions of consciousness more directly in view, designing experiments with the CBA explicitly in mind.

7 Conclusion

The question of what function(s) consciousness has, if any, has an important bearing on many debates that are central to our understanding of consciousness. It promises to inform accounts of how and why consciousness evolved (Nichols and Grantham 2000); guide attempts to identify the distribution of consciousness in animals and machines (Allen and Trestman 2024; Bayne et al. 2024); and bears on the question of whether consciousness is a natural kind (Bayne and Shea 2020; Irvine 2012).

Rather than attempting to say what the function(s) of consciousness are, this paper has tackled the more foundational question of how an account of the function(s) of consciousness ought to be structured. We have argued that an account of the function(s) of consciousness ought to be structured in terms of capacities. This capacity-based approach to the function(s) of consciousness, we have argued, is grounded in a well-motivated approach to the explanation of the mind generally, avoids problems and mistakes that hinder other approaches to the function(s) of consciousness, and re-orients explanatory priorities in ways that suggest novel, fruitful lines of experimentation. We think it should be used to guide ongoing research in consciousness science.

Acknowledgements

Thanks to two referees for this journal for their very helpful comments on earlier versions of this paper. Thanks also to the audience at the 2023 meeting of the European Society for Philosophy and Psychology, and to colleagues in the ‘Brain, Mind, and Consciousness’ program of the

Canadian Institute for Advanced Research (CIFAR), especially Bob Kentridge, for multiple discussions.

Funding

TB and JS gratefully acknowledge support from CIFAR catalyst grant ‘Functions of Consciousness’; JS gratefully acknowledges support ERC Starting Grant 757698, and from the Templeton World Charity Foundation, Inc. (funder DOI 501100011730) under the grant https://doi.org/10.54224/32583.

ORCID

Joshua Shepherd: https://orcid.org/0000-0003-2008-649X

Tim Bayne: https://orcid.org/0000-0001-8591-7907

References

Allen, Colin and Michael Trestman (2024) ‘Animal consciousness’, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta and Uri Nodelman (eds.), <https://plato.stanford.edu/archives/sum2024/entries/consciousness-animal/>.

Anderson, Michael L. (2007) ‘Massive redeployment, exaptation, and the functional integration of cognitive operations’, *Synthese* 159(3): 329-345.

Amihai, Ido, Leon Deouell, and Shlomo Bentin (2011) ‘Conscious awareness is necessary for processing race and gender information from faces’, *Consciousness and cognition* 20(2): 269-279.

Baars, Bernard J. (1993) *A Cognitive Theory of Consciousness*. Cambridge University Press.

Baars, Bernard J. (2005) 'Global workspace theory of consciousness: toward a cognitive neuroscience of human experience', *Progress in Brain Research* 150: 45-53.

Bayne, Tim (2008) 'The phenomenology of agency', *Philosophy Compass* 3: 1-21.

Bayne, Tim (2010) *The Unity of Consciousness*. Oxford University Press.

Bayne, Tim, Jakob Hohwy, and Adrian M. Owen (2016) 'Are there levels of consciousness?' *Trends in Cognitive Sciences* 20(6): 405-413.

Bayne, Tim, and Jakob Hohwy (2016) 'Modes of consciousness', In Walter Sinnott-Armstrong, ed., *Finding Consciousness: The Neuroscience, Ethics and Law of Severe Brain Damage*: 57-80. Oxford University Press.

Bayne, Tim, Anil K. Seth, Marcello Massimini, Joshua Shepherd, Axel Cleeremans, Stephen M. Fleming, Rafael Malach, Jason B. Mattingley, David K. Menon, Adrian M. Owen, and Megan A. Peters (2024) 'Tests for consciousness in humans and beyond', *Trends in Cognitive Sciences* 28(5): 454-466.

Bayne, Tim, and Nicholas Shea (2020) 'Consciousness, concept, and natural kinds', *Philosophical Topics* 48(1, 65-84.

Block, Ned (2016) 'The Anna Karenina principle and skepticism about unconscious perception', *Philosophy and Phenomenological Research* 93(2): 452-459.

Block, Ned (2023) *The Border between Seeing and Thinking*. Oxford University Press.

Boly, Melanie, Robert D. Sanders, George A. Mashour, and Steven Laureys (2013) 'Consciousness and responsiveness: lessons from anaesthesia and the vegetative state', *Current Opinion Anaesthesiology* 26: 444-449.

Carey, Susan (2009) 'Where our number concepts come from', *The Journal of Philosophy* 106(4): 220-254.

Casali, Adenauer G., Olivia Gosseries, Mario Rosanova, Mélanie Boly, Simone Sarasso, Karina R. Casali, Silvia Casarotto, Marie-Aurélie Bruno, Stephen Laureys, Giulio Tononi, and Marcello Massimini (2013) 'A theoretically based index of consciousness independent of sensory processing and behavior', *Science Translational Medicine* 5(198): 198ra105.

Charles, Lucie, Filip Van Opstal, Sébastien Marti, and Stanislas Dehaene (2013) 'Distinct brain mechanisms for conscious versus subliminal error detection', *Neuroimage* 73: 80-94.

Christensen, Wayne, Sutton, John, and Doris J. McIlwain (2016) 'Cognition in skilled action: Meshed control and the varieties of skill experience', *Mind & Language* 31(1): 37-66.

Craver, Carl F. (2001) 'Role functions, mechanisms, and hierarchy', *Philosophy of Science* 68(1): 53-74.

Craver, Carl F. (2007) *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Clarendon Press.

Cummins, Robert (1975) 'Functional analysis', *Journal of Philosophy* 72: 741–765.

Dehaene, Stanislas, Michel Kerszberg, and Jean-Pierre Changeux (1998) 'A neuronal model of a global workspace in effortful cognitive tasks', *Proceedings of the National Academy of Sciences* 95(24): 14529-14534.

Desender, Kobe, and Eva Van den Bussche (2012) 'Is consciousness necessary for conflict adaptation? A state of the art' *Frontiers in human neuroscience* 6: 3.

Dijksterhuis, Ap, Maarten W. Bos, Loran F. Nordgren, and Rick B. Van Baaren (2006) 'On making the right choice: The deliberation-without-attention effect' *Science* 311(5763): 1005-1007.

Edelman, Gerald M., and Giulio Tononi (2000) *A Universe of Consciousness: How Matter Becomes Imagination*. Basic Books.

Firestone, Chaz, and Brian J. Scholl (2016) 'Cognition does not affect perception: Evaluating the evidence for 'top-down' effects', *Behavioral and Brain Sciences* 39: E229.

Francken, Jolien C., Lola Beerendonk, Dylan Molenaar, Johannes J. Fahrenfort, Julian D. Kiverstein, Anil K. Seth, and Simon Van Gaal (2022) 'An academic survey on theoretical foundations, common assumptions and the current state of consciousness science', *Neuroscience of Consciousness* 1: niac011.

Godfrey-Smith, Peter (1994) 'A modern history theory of functions', *Noûs* 28(3): 344-362.

Irvine, Elizabeth (2012). *Consciousness as a Scientific Concept*. Springer.

Irvine, Elizabeth, and Mark Sprevak (2020) 'Eliminativism about consciousness', In Uriah Kriegel, ed., *The Oxford Handbook for the Philosophy of Consciousness*: 348-370. Oxford University Press.

Keizer, André W., Bernhard Hommel, and Victor A. F. Lamme (2015) 'Consciousness is not necessary for visual feature binding', *Psychonomic Bulletin & Review* 22: 453-460.

Kentridge, Robert W., Charles A. Heywood, and Lawrence Weiskrantz (1999) 'Attention without awareness in blindsight', *Proceedings of the Royal Society of London, B* 266(1430)*:* 1805-11.

Klein, Colin, Jakob Hohwy, and Tim Bayne (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).

Krickel, Beate (2018) 'A regularist approach to mechanistic type-level explanation', *The British Journal for the Philosophy of Science* 69(4): 1123-1153.

Kriegel, Uriah (2015) *The Varieties of Consciousness*. Oxford University Press.

Lamme, Victor A. (2006) 'Towards a true neural stance on consciousness', *Trends in Cognitive Sciences* 10(11): 494-501.

Lassiter, G. Daniel, Matthew J. Lindberg, Claudia González-Vallejo, Francis S. Bellezza, and Nathaniel D. Phillips (2009) 'The deliberation-without-attention effect: Evidence for an artifactual interpretation', *Psychological Science* 20(6): 671-675.

Laureys, Stephen (2005) 'The neural correlate of (un)awareness: lessons from vegetative state', *Trends in Cognitive Sciences* 9: 556-559

Libet, Benjamin (1985) 'Unconscious cerebral initiative and the role of conscious will in voluntary action' *Behavioral and Brain Sciences* 8(4): 529-539.

Ludwig, Dylan (2022) 'The functional contributions of consciousness', *Consciousness and Cognition* 104: 103383.

McDougle, Samuel D., Richard B. Ivry, and Jordan A. Taylor (2016) 'Taking aim at the cognitive side of learning in sensorimotor adaptation tasks', *Trends in Cognitive Sciences* 20(7): 535-544.

Machamer, Peter, Lindley Darden, and Carl F. Craver (2000) 'Thinking about mechanisms', *Philosophy of Science* 67(1): 1-25.

McKilliam, Andrew (2020) 'What is a global state of consciousness?' *Philosophy and the Mind Sciences* 1(II).

Michel, Matthias (2022) 'Confidence in consciousness research', *Wiley Interdisciplinary Reviews: Cognitive Science* 14(2): e1628.

Morsella, Ezequiel (2005) 'The function of phenomenal states: Supramodular interaction theory', *Psychological Review* 112(4): 1000-1021.

Mudrik, Liad, Nathan Faivre, and Christof Koch (2014) 'Information integration without awareness', *Trends in Cognitive Sciences* 18(9): 488-496.

Murray, D. J. (1968) 'Articulation and acoustic confusability in short-term memory', *Journal of Experimental Psychology* 78: 679.

Nguyen, Cong Tam, Gilles Roy, Christian Gauthier, and Nicolas Galani (2007) 'Heat transfer enhancement using Al2O3–water nanofluid for an electronic liquid cooling system', *Applied Thermal Engineering* *27*(8-9): 1501-1506.

Nichols, Shaun, and Todd Grantham (2000) 'Adaptive complexity and phenomenal consciousness', *Philosophy of Science* 67(4): 648-670.

Niikawa, Takuya, Katsunori Miyahara, Hiro Taiyo Hamada, and Satoshi Nishida (2022) 'Functions of consciousness: conceptual clarification', *Neuroscience of Consciousness* 1: niac006.

Overgaard, Morten, and Rikke Overgaard (2010) 'Neural correlates of contents and levels of consciousness', *Frontiers in Psychology* 1(164): 1-3.

Pacherie, Elizabeth (2008) 'The phenomenology of action: A conceptual framework', *Cognition* 107(1): 179-217.

Persuh, Marjan, and Alexander Rue (2019) 'Unconscious working memory outside the focus of attention', *Journal of Vision* 19(10): 75b-75b.

Peters, Megan A., Robert W. Kentridge, Ian Phillips, and Ned Block (2017) 'Does unconscious perception really exist? Continuing the ASSC20 debate', *Neuroscience of Consciousness* 3(1).

Phillips, Ian (2018) 'Unconscious perception reconsidered', *Analytic Philosophy* 59(4): 471-514.

Piccinini, Gualtiero (2020) *Neurocognitive mechanisms: Explaining biological cognition*. Oxford University Press.

Piccinini, Gualtiero, and Carl Craver (2011) 'Integrating psychology and neuroscience: Functional analyses as mechanism sketches', *Synthese* 183: 283-311.

Povich, Mark (2015) 'Mechanisms and model-based functional magnetic resonance imaging', *Philosophy of Science* 82(5): 1035-1046.

Povich, Mark, and Carl F. Craver (2017) 'Mechanistic levels, reduction, and emergence', In Stuart Glennan and Phyllis Illari, eds., *The Routledge Handbook of Mechanisms and Mechanical Philosophy*: 185-197. Routledge.

Robinson, Zack, Corey J. Maley, and Gualtiero Piccinini (2015) 'Is consciousness a spandrel?' *Journal of the American Philosophical Association* 1(2): 365-383.

Seth, Anil (2009) 'Functions of consciousness', In William P. Banks, ed., *Encyclopedia of Consciousness*: 279-293. Elsevier.

Shallice, Tim (1972) 'Dual functions of consciousness', *Psychological Review* 79(5): 383-393.

Shepherd, Gordon M. (1994) *Neurobiology (3rd ed.)*. Oxford University Press.

Shepherd, Joshua (2015) 'Conscious control over action', *Mind & Language* 30(3): 320-344.

Shepherd, Joshua (2016) 'Conscious action/zombie action', *Noûs* 50(2): 419-444.

Shepherd, Joshua, and Myrto Mylopoulos (2021) 'Unconscious perception and central coordinating agency', *Philosophical Studies* 178(12): 3869-3893.

Siegel, Susanna (2012) 'Cognitive Penetrability and Perceptual Justification', *Noûs* 46(2): 201-222.

Silverman, Michael E., and Arien Mack (2006) 'Change blindness and priming: When it does and does not occur', *Consciousness and Cognition* 15(2): 409-422.

Stein, Timo, and Marius V. Peelen (2021) 'Dissociating conscious and unconscious influences on visual detection effects', *Nature Human Behaviour* 5(5): 612-624.

Triggiani, Antonio I., Gabriel Kreiman, Cara Lewis, Uri Maoz, Alfred Mele, Liad Mudrik, Adina L. Roskies, Aaron Schurger, and Mark Hallett (2023) 'What is the Intention to Move and When Does it Occur?' *Neuroscience & Biobehavioral Reviews* 151: 105199.

Tye, Michael (1996) 'The function of consciousness', *Noûs* 30(3): 287-305.

Webb, Taylor W., Hope H. Kean, and Michael S. A. Graziano (2016) 'Effects of awareness on the control of attention', *Journal of Cognitive Neuroscience* 28(6): 842-851.

Wegner, Daniel M. (2004) 'Précis of the illusion of conscious will', *Behavioral and Brain Sciences* 27(5): 649-659.

Weiskopf, Daniel A. (2011) 'Models and mechanisms in psychological explanation', *Synthese* 183: 313-338.

Wright, Larry (1973) 'Functions', *The Philosophical Review* 82(2): 139-168.