



Predictive processing as a systematic basis for identifying the neural correlates of consciousness

Jakob Hohwy^a  (Jakob.Hohwy@monash.edu)

Anil Seth^b  (a.k.seth@sussex.ac.uk)

Abstract

The search for the neural correlates of consciousness is in need of a systematic, principled foundation that can endow putative neural correlates with greater predictive and explanatory value. Here, we propose the predictive processing framework for brain function as a promising candidate for providing this systematic foundation. The proposal is motivated by that framework's ability to address three general challenges to identifying the neural correlates of consciousness, and to satisfy two constraints common to many theories of consciousness. Implementing the search for neural correlates of consciousness through the lens of predictive processing delivers strong potential for predictive and explanatory value through detailed, systematic mappings between neural substrates and phenomenological structure. We conclude that the predictive processing framework, precisely because it at the outset is not itself a theory of consciousness, has significant potential for advancing the neuroscience of consciousness.

Keywords

Active inference · Consciousness · Neural correlates · Prediction error minimization

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

1 Introduction

The search for the *neural correlates of consciousness* (NCCs) has now been underway for several decades (Crick & Koch, 1990b). Several interesting and suggestive

^aCognition & Philosophy Lab, Monash University

^bSackler Center for Consciousness Science and School of Engineering and Informatics, University of Sussex; Azrieli Program in Brain, Mind, and Consciousness Toronto

patterns of brain activity have been revealed but no unequivocal ‘correlate’ has emerged. Alongside these developments, a plethora of *theories* of consciousness have been formulated. Many of these theories are conceptually intriguing, and some have empirical evidence in their favour, but the field of consciousness science is some distance from consensus on a theory of consciousness (for overviews, see Bayne et al., 2009); the lack of consensus has sparked funding bodies to encourage adversarial collaborations between proponents of different theories (Reardon, 2019). As with NCCs, this lack of consensus can be attributed partly to continuing debate over the sense in which an adequate theory should ‘explain’ ‘consciousness’. There is a further distinction between scientific and metaphysical theories of consciousness; the former typically at least tacitly assumes materialism and aims at explanation and mechanistic approaches to consciousness, whereas the latter is concerned with the ultimate nature of consciousness rather than with specifics about neuronal mechanisms.

Here, we propose that it will be fruitful for the field to capitalize on a general account of brain function, which can serve as a systematic basis for identifying NCCs, and which may as a consequence increase the predictive and explanatory power of the science of consciousness. We suggest, in particular, that a predictive processing (PP) approach brings many advantages to the scientific investigation of consciousness, primarily through encouraging and operationalizing closer links between phenomenological properties of conscious experience and mechanistic properties of underlying neural substrates. Perhaps counter-intuitively, we suggest these advantages accrue in part because the PP approach is not, in itself, a theory of consciousness.

2 The need for a systematic NCC

To identify the neural correlates of consciousness, it is commonly accepted that one needs to find the minimally sufficient conditions for the occurrence of a global conscious state or level (such as being conscious rather than unconscious, or being in the Minimally Conscious State rather than the awake state) or content (such as consciously experiencing a face rather than a house). There exist a number of different methods and typical paradigms for revealing such correlates. For example, binocular rivalry and masking paradigms have been extensively used in studies of conscious content, while for the global conscious state the natural or artificial manipulations of (respectively) sleep and anaesthesia provide productive approaches.

These NCC methods are generally, at least implicitly, considered relatively theory neutral – they are not immediately proprietary to any particular scientific theory of consciousness (and proponents of most metaphysical theories, even dualists, can agree that these methods implement a valid scientific endeavor even if, by their lights, they cannot reveal the ultimate truth about consciousness). There is discussion about whether states can be conscious even if not accessed and reportable, and if that is the case, whether the neural correlates of such states can be identi-

fied by paradigms that require behavioural report (Block, 1995). There has been progress in extending NCC methods across conditions where introspective or behavioural report are not required and such ‘no report’ paradigms may begin to address this discussion (Tsuchiya et al., 2015). While NCC methods may initially appear to be theory-neutral, different theories may gravitate towards certain types of paradigms and proposed neural correlates. For example, theories that deny the possibility of unaccessed conscious experience will tend to gravitate to paradigms that capitalise on reportability.

After paradigm-defining research in the early 1990s, the NCC methodology became a focus of the science of consciousness (Metzinger, 2000), and has received several reviews and commentaries since, reporting on significant progress (Aru et al., 2012; Fink, 2016; Hohwy, 2009; Klein & Barron, 2020; Koch et al., 2016; Miller, 2007; Seth, 2009; Tononi & Koch, 2008). An influential early discussion was given in Chalmers (Chalmers, 2000), which pointed to a crucial and so far underappreciated distinction between *arbitrary* and *systematic* NCCs.

An arbitrary NCC is revealed when some neural activity in some region of the organism’s brain (as accessed with some neuroimaging or neurophysiological method) reliably correlates with a given conscious state or content. Such a correlate is ‘arbitrary’ because mere activation on its own does not help us extrapolate, predict, understand or explain much about the mechanisms underlying consciousness. The fact that region x ‘lights up’ when consciously rather than unconsciously perceiving Gabor patches does not, by virtue of this fact alone, allow us to predict under what other stimulus conditions x might also be activated, or which regions might be activated for other types of stimuli. Nor does it throw much light on the particular phenomenological properties that characterize such a conscious perception, nor what mechanisms underlie conscious perception generally. Even with extensive brain mapping along these lines, the result would be an explanatorily inert descriptive list of areas or patterns of activity. Arbitrary NCCs are not likely to do much to let us understand the brain basis of consciousness, or to use such an understanding to help with consciousness-related conditions, such as disorders of consciousness (Fernández-Espejo & Owen, 2013), split brains (De Haan et al., 2020), or islands of consciousness (Bayne et al., 2020).

As a terminological note, we take ‘arbitrary’ in the sense that we would have no inkling why the pattern of activity would be like this and not some other way, given some particular phenomenology; such a fully arbitrary NCC is one extreme kind of NCC. We recognize that talk of ‘arbitrary’ NCCs may not be entirely felicitous since the NCC in question may have a reliable correlation with phenomenal states that is law-like and exceptionless – and in this sense be ‘non-arbitrary’. In some respects, a better distinction could be between ‘non-explanatory’ and ‘explanatory’ correlates, as in (Seth, 2009). Note however that the arbitrariness we (following Chalmers, 2000) refer to is not in the correlation, reliable or not, with phenomenal states, but rather in the question why it is this rather than another correlated pattern of neural activity that is the focus of the NCC search. Being

‘systematic’ then means being systematically guided by theoretical considerations of some sort.

Most actual instances of NCC research will not be fully arbitrary, as they naturally engage to some degree interpretations of the pattern or location of activation, mainly using that pattern or location to inspire a theoretical approach to consciousness. Examples of such theory building can be found in the idea that frontoparietal activity is not arbitrary in the sense that it accounts for the global broadcast and access that seems predictive and explanatory at least for access consciousness (Mashour et al., 2020), or in the interpretation of the role of ‘posterior hubs’, associated with integrated information theory approaches (Koch et al., 2016).

In light of such concerns, Chalmers notes that if we could “find a neural system whose states do not just arbitrarily correlate with the phenomenal [contents and] states in question, but which vary along a corresponding *systematic* dimension, then the NCC in question will have much greater potential explanatory and predictive power”. We agree with this reasoning and label this the *systematicity requirement* for the NCC. (Chalmers suggested as an example of systematicity correlations of conscious content with ‘neural representational systems’ (Chalmers, 2000: 11); for critical discussion, see Noë & Thompson (2004) and responses; in this paper, we will consider a more inclusive notion than just representational systems).

To unpack the systematicity requirement, we must first determine with respect to what there should be greater explanatory and predictive power. The answer to this question is that explanatory and predictive power should be with respect to both *brain activity* and *phenomenology*. A systematic approach will be useful for the NCC search only if its core elements can eventually be used to predict when and how conscious phenomena arise, explain why conscious phenomena have the structure and form they have, and potentially even control through interventions how conscious experience unfolds.

We propose that such a systematic approach will best be developed by embedding the proposed correlates of the studied conscious phenomena within a larger understanding of brain function, one which is not itself primarily tied to consciousness. This will help move beyond mere arbitrary correlates by ensuring that the relevant brain activity can be identified and interpreted as more than just “the activity, whatever it is, that correlates with this or that conscious phenomenon”. In other words, it would enable us to look at the correlated brain activity and project systematically from that activity to predict related phenomenological states, and provide unifying explanations of their occurrence.

We can then say that, to discharge the systematicity requirement, a larger understanding, theory or, as we shall phrase it, a *framework* for overall brain function must be found for the NCC. This framework must perform the task of providing explanatory and predictive power for certain patterns of brain activity *and* the structure and form of their correlated phenomenology. Note, it is not necessary for such a framework to itself propose an explanation of consciousness. In fact, it

may serve the search for a systematic NCC better if this larger framework does not – at least not initially – contain a theory of consciousness. As we will discuss later, parts of the framework may nevertheless coalesce into a theory of consciousness if it turns out to be successful in contributing to a systematic NCC. (We do not argue that discharging the systematicity requirement necessarily requires a framework that is not a theory of consciousness. For example, Integrated Information Theory is a theory of consciousness that proposes systematic relationships between neural mechanisms and phenomenological properties (Tononi et al., 2016). Note that IIT is not generally considered as part of the NCC programme.)

Naturally, it will be a tremendous challenge to fully cash out this notion of a framework for overall brain function with the requisite explanatory and predictive power. We believe, as we shall discuss later, that there is cause for optimism. For now, an analogy might be useful, that the systematicity requirement is not specific to consciousness science. If we are interested in physiological phenotypic traits then a basic genetic theory of what governs changes to the DNA sequence will not be useful on its own, even if that theory is in fact true. Rather, to gain explanatory and predictive power with respect to phenotype we need to appeal to epigenetics. Epigenetics places genes in context and allows us to build up mechanistic accounts of how genes, cells, bodies, and environments interact over developmental trajectories to explain phenotypic traits. Epigenetics is a theoretical approach in its own right and there is considerable work to be done in correlating epigenetic processes with phenomena like cancer. Similarly, for a theory of brain function, there will be kinds of frameworks that are not optimal for generating explanatory and predictive power for consciousness, even if they might adequately describe (for example) neuronal and synaptic activities, while there will be frameworks at other levels of description that will have the right kind of explanatory and predictive ‘grain’ for consciousness.

With the help of a suitable framework for brain function, the fact that conscious perception of a Gabor patch correlates with brain activity x would have greater potential for predicting under what other conditions x would or wouldn’t be present, and under what conditions a similar conscious perception would or wouldn’t be present. Critically, such a framework should be capable of providing an adequate language in which to account for phenomenological properties of consciousness (e.g., representational content, sensory presence, feature binding, spatiality or the lack of spatiality, illusion formation, sense of agency etc.), which constitute the explananda of any satisfactory theory of consciousness. That is, unpacking brain activity x within the framework should cast light on phenomenology.

In this manner, it would be useful for the empirical search for NCCs to be coupled to a framework, not for consciousness per se but rather a framework for brain function, with the right internal structure and level of description to have explanatory potential for phenomenology. Such an approach contrasts with the NCC search as it has largely unfolded until now, which may be characterized as identifying relatively arbitrary and piecemeal correlates.

3 Three challenges to the NCC search

Identifying the general framework for brain function for the systematic NCC will be helped by considering three additional and extensively discussed challenges associated with the NCC search.

3.1 The global states and local contents challenge

The first challenge has to do with the widely acknowledged distinction between *states* and *contents* of consciousness (Laureys, 2005; Rosenthal, 1986). Most NCC programs are devoted to content NCCs, and fewer (though increasingly more) to finding the NCC of global states (sometimes called ‘levels’) of consciousness. The challenge here is that a satisfying account of consciousness may require treating states and contents together (Bayne, 2007; Hohwy, 2009). The challenge can best be introduced with an example. Binocular rivalry is a classic NCC paradigm, where different stimuli are shown continuously to each eye, and conscious perception alternates between them. In a typical binocular rivalry NCC study, a participant may be exposed to rivalry stimuli and fMRI data may be acquired, showing alternating regions of activity associated with the alternating conscious percepts (as in, e.g., the classic study by Tong et al., 1998). The problem, stressed early on by Searle (2000), is that the participant was already conscious when put in the scanner and so the brain activity cannot reveal why one percept is conscious rather than not, but merely what brain activity is associated with selection of the content into consciousness, in an *already* conscious creature.

More generally, the NCC search has largely sidestepped the issue of the relation between global states and contents (exceptions include Bachmann, 2012; Aru et al., 2019; Haun & Tononi, 2019; Noreika et al., 2017). Much remains to be discovered about interactions between global states and contents. For example, it may be that insertion of some contents into some global states changes those states, or that certain states may change the contents inserted into them (e.g., rivalry stimuli may be processed very differently in different global states of consciousness such as delirium, psychedelic states, or absence seizures). Indeed, the psychedelic state serves as a compelling example in which there seem to be changes in both state/level and content, which, further, seem to be dependent on each other (Schartner et al., 2017).

This challenge constrains the search for a systematic NCC in the following way: the systematic dimension under which neural activity is described should be able to have both global states and local content as factors. There should be dimensions of overall functioning – a ‘whole of system’ approach – as well as of individual conscious contents.

3.2 The screening off challenge: Brain structure and phenomenology

The second challenge concerns how to ‘screen off’ less explanatory neural correlates from the more explanatory correlates. Importantly, the degree to which a candidate NCC is explanatory is not simply a matter of how tight or reliable its correlation is with conscious states or contents. A number of researchers, beginning with Miller (2007), have argued that even with very tight, reliable correlates being identified, it may still be an open question which of these are the true constituents or substrates of the conscious states or contents (with the underlying assumption being that only the true constituents would facilitate predictive and explanatory power). If there are three causally chained neural states NS1, NS2, NS3 that tightly and reliably correlate with the conscious experience of seeing a face, it may be that NS1 is a mere causal precursor, and NS3 a causal consequent of the true (and presumably most explanatory) constituent NS2, even though the empirical correlation between all three and the conscious experience is much the same (Graaf et al., 2012; Miller, 2015).

One response to the screening off challenge is to make use of techniques for causal inference (Aru et al., 2012; Hohwy, 2009; Klein & Barron, 2020; Neisser, 2012; Pearl, 2000; Woodward, 2003). In particular, better methods for causal *intervention* should be able to advance the issues. For example, TMS, tDCS, and especially optogenetics should allow fine-grained (‘surgical’) manipulation of distinct elements of the causal neural chains of events underlying conscious states and contents (see Marshel et al. (2019) for an impressive optogenetic study along these lines). This holds most obviously for precursors, where interventions would lead to effects on consciousness. It also holds for downstream consequences, where interventions should fail to influence consciousness.

A second and perhaps more important response lies in examining whether potential correlates (e.g., NS2) better account for the corresponding conscious content or state than precursors (NS1) or consequents (NS3). This requires the ability to distinguish potential correlates on their explanatory potential with (phenomenological) explanatory targets. For example, NS2 and NS1 may correlate equally tightly with a conscious experience of a horizontal Gabor patch, but NS1 correlates better than NS2 with several orientations of the experienced patch. Here, it is a question for philosophy of science to formulate a substantial notion of ‘explanatory potential’; this may be, for example, in terms of homology, causal/functional isomorphism, contrastive explanations, or mechanistic reduction (Bechtel, 2007; Craver, 2007; Hohwy & Frith, 2004; O’Brien & Opie, 1999).

Meeting the screening off challenge would therefore be facilitated if brain activity could be ‘carved at the joints’ in a way that both allowed meaningful and efficient interventions and which mapped to an equivalent partitioning of the explanatory targets, namely the phenomenological properties of experience. There is a considerable and challenging task in revealing this ‘architecture’ of conscious

phenomenology, and it seems likely that approaches to screening off at the level of brain activity will develop in tandem with understanding of the phenomenological explanatory targets. In particular, we may learn more about the building blocks of the neural system and the varieties of phenomenology as we learn to better intervene on both.

A standard example serves to illustrate the difficulties inherent in carving brain activity at its joints such that it matches the structure of the phenomenological explanatory targets. An early dynamical NCC approach proposed that conscious contents depended on gamma-band synchrony among neural populations (Crick & Koch, 1990a, 1990b). This proposal gained conceptual plausibility through its association with the classic ‘binding problem’ in neuroscience of how spatially segregated processing of distinct object features become ‘bound’ together. Despite some early empirical support it became rapidly clear that synchrony among neural populations does not always go along with conscious contents and/or binding (for review and discussion, see Gold, 1999; Engel & Singer, 2001). To the contrary, an excess of neuronal synchrony reliably leads to the global loss of consciousness, as in absence epilepsy. In hindsight, the insufficiency of a simple association between neural synchrony and consciousness is hardly surprising. Allowing a more global view of binding, conscious scenes are not well described as single unified perceptual objects; rather, conscious experiences normally comprise of a multi-modal range of contents each of which retain a distinctive identity while being embedded within a single unified scene. This phenomenological observation may then be a good guide for how to carve brain activity at the joints, where neuronal synchrony may still play some role – though no longer as the NCC itself (Tononi & Edelman, 1998).

3.3 The action challenge

We turn next to a challenge which is more prescriptive, about where consciousness science should go, than the previous challenges. Recent years have seen an increased focus on the role of *action* (in a broad and inclusive sense) for the mind. A number of broadly ‘enactive’ approaches have been suggested, which highlight the important point that conscious perception is closely entwined with agency, intention, movement and behaviour (contributions to this debate range very widely from radical reconceptualizations of cognitive science, biological approaches to active sensing, and computational approaches, see, e.g., Hurley, 1998; Engel et al., 2016; Hutto & Myin, 2013; Krakauer et al., 2017; Noë, 2004; Noë & O’Regan, 2001; Schroeder et al., 2010). Indeed, some of the pressure to include action in consciousness science comes from the predictive processing literature itself, insofar as this literature rests on a notion of action (as we discuss later) and has already begun to address consciousness (Parr et al., 2019), and so insisting on this challenge will somewhat preemptively help secure a good position for this framework.

Our conscious experience is intimately coupled with our actions at both personal levels (for example, a voluntary decision to reach for a cup of tea) and sub-personal levels (for example, as we actively sample the world in visual search, by moving our eyes around a visual scene). Though we might readily imagine an utterly actionless but conscious system, it seems misguided to suggest that action is an optional extra for any actual system that is conscious. Notice, this is not just the idea that consciousness is *for* action, but rather that the conscious experience has the structure and trajectory it has in part because we are active systems.

Presently, neuroscientific theories of consciousness and NCC research that focus on conscious states and contents pay little attention to the role of action in consciousness. The natural exceptions are studies that explicitly focus on experiences of intention, volition and agency, but these studies are rarely taken to motivate general accounts of consciousness or brain function (see also Fleming et al. (2015) for an intriguing link between action and metacognition). One possible reason for general neglect of action is that action and consciousness seem to be dissociable, at least on the face of it. For example, we can be conscious without overt action as in dreaming, and we can act without consciousness, as in reflexes. Nevertheless, a systematic dimension to brain function that includes action as an integrated part would be desirable, allowing us to posit the necessity of action for consciousness. Such proposals might include covert, mental actions (such as endogenous allocation of attention) as well as overt actions – and they might distinguish between actions being constitutively necessary (i.e., necessary at any given time for consciousness), compared to being merely diachronously necessary (i.e., necessary as part of the history of the organism’s interaction with its environment).

Summarizing these three challenges for the systematic NCC search, we would like to have a framework for interpreting brain activity that allows interactions between global and local states of consciousness, enables intervention at ‘joints’ which make sense both in terms of brain activity and phenomenology, and which allows an integrated role for action in consciousness. Next, after considering some theoretically motivated constraints, we will propose that the predictive processing framework fits the bill.

4 Theories of consciousness

There are now many theories of consciousness vying for attention and empirical evidence. Here is an incomplete list of some of the most discussed theories, with some exemplifying references:

Global neuronal workspace theory (Baars, 1988; Mashour et al., 2020)

Integrated information theory (Tononi et al., 2016; Tononi & Edelman, 1998)

Recurrent processing theory (Lamme, 2020, 2010)

Higher order thought theories & Metacognitive theories (Brown et al.,

2019; Fleming, 2020; Lau, 2008; Lau & Rosenthal, 2011; Rosenthal, 1997)

Radical plasticity thesis (Cleeremans, 2011)

Virtual reality theories (Metzinger, 2004; Revonsuo, 2006)

Attention-based theories (Graziano & Webb, 2015; Prinz, 2012)

Heterophenomenology (Dennett, 1991)

Core consciousness theory (Damasio, 2000)

Orchestrated objective reduction (Hameroff & Penrose, 2014)

Electromagnetic theory (McFadden, 2020, 2002)

Some of these labels and groupings ignore substantial differences. For example, though versions of Higher-order thought theory are closely linked to Metacognitive theories they differ significantly in how they cash out the higher-order element; the two Attention-based theories differ, though both appeal to attention, and similarly for the two Virtual reality theories. There are also now several theories proposed which connect to the framework we shall consider below, namely predictive processing; we discuss these theories below. Though there is some diversity in the explanatory targets and in metaphysical assumptions (for discussion of this kind of diversity, see e.g., Seth, 2021), there is general agreement that a theory of consciousness should explain the underlying mechanism(s) for conscious contents and/or global states of consciousness.

Regarding metaphysical assumptions, there is some debate about whether theories of consciousness are or should be aimed at resolving the metaphysical mind-body problem (the “hard problem”, see also alternate versions such as the knowledge argument or the explanatory gap; Descartes, 1641; Chalmers, 1996; Jackson, 1982; Levine, 1983). While some theories explicitly seek to solve the hard problem (e.g., Integrated information theory, Orchestrated objective reduction, Electromagnetic theory), others do not, or do so only implicitly or indirectly (e.g. recurrent processing theory, global neuronal workspace theory). To solve the hard problem, a theory must remove any open question about whether a system characterised in terms of that theory is in fact conscious. This is a tall order. It seems to us the best hope for making inroads on the metaphysical question is to take an indirect approach, and to extend this approach by searching for *systematic* neural correlates of consciousness (though we acknowledge of course that the hard problem may be too metaphysically hard for any scientific approach to be successful). On one level, this strategy will deliver greater predictive and explanatory power regarding factors that shape and modulate consciousness. Such advances will have important practical consequences in (for example) neurology and psychiatry. Equally important, a focus on systematic NCCs recognizes that ‘consciousness’ is not a singular explanatory target. The multifaceted structure of conscious phenomenology instead provides a repertoire of explanatory targets. Applying a general theory of brain function alongside a refined repertoire of explanatory targets (as opposed to implicitly assuming consciousness to be a singular phenomenon) should prove powerful in circumventing the hard problem of why conscious experience happens

at all (Seth, 2016). In other words, a sensible mapping – achieved by satisfying the systematicity constraint – of brain states and phenomenology may make the hard problem less pressing, or perhaps even dissolve it altogether.

As mentioned, later we examine extensively whether predictive processing can provide the theoretical framework needed to deliver a systematic NCC. This discussion will raise the issue whether PP is itself a theory of consciousness (and really belongs on the list above), or is on the way to becoming one. For now, we note that there have been several recent proposals for how to connect PP to consciousness. Our point here is that PP is not first and foremost a theory of consciousness since, as a framework for brain function, it could exist perfectly well even in the absence of any discussion of consciousness. This fact about it will, as we shall see, encourage its consideration as the basis for a systematic NCC.

4.1 Common themes

In spite of the diversity of current theories of consciousness, the most prominent have at least two elements in common. If these theories are at least roughly on the right track, then it will be desirable if the underlying framework for a systematic NCC, which we seek, can give rise to these common elements.

The first element is *uncertainty reduction*. Many theories associate consciousness with a mechanism such that experiencing a certain content in a certain conscious state is related to uncertainty reduction, in an information-theoretical sense. For example, Global neuronal workspace theory operates with the idea of contents being ignited into consciousness when an evidence accumulation threshold is reached (Dehaene, 2011); versions of Higher-order thought theory operate with criterion setting that optimizes a variable signal-to-noise ratio (Lau, 2008), the same holds for the Radical plasticity thesis, and Metacognitive theories; Integrated information theory formally defines conscious states in part in terms of the possibilities excluded by being in that state (Oizumi et al., 2014); Recurrent processing theory considers whether the mechanisms for consciousness have to do with integrating information and with learning, both of which are uncertainty reducing. The reasoning behind this insistence on uncertainty reduction often seems to be that a conscious system has a tendency to settle in one unified and highly informative representational state. Of course, we are not here placing all these diverse theories on the same footing with respect to how they conceive such uncertainty reduction. Different theories propose diverse and often incompatible mechanisms, operating at quite different spatiotemporal and even conceptual scales (for discussion, see Doerig et al., 2020; Graziano et al., 2020).

The second element that many theories of consciousness have in common is *top-down signalling*. Many theories of consciousness agree that there is little chance for consciousness to occur in a system that does not have top-down ('backward' or 'inside-out') signalling in addition to the bottom-up ('forward' or 'outside-in') flow from the sensory cortices. For Integrated information theory, it turns out

that a system with only bottom-up connectivity has no ϕ ('phi', the quantitative measure of consciousness on this theory; (Oizumi et al., 2014: 19); Recurrent processing theory is based on the notion of backward, 'recurrent' processing; NCC studies in support of Global neuronal workspace theory suggest that consciousness is 'ignited' partly by recurrent connectivity; Higher-order thought and Metacognitive theories all operate with an assumption that some higher-level process can 'look down' at lower-level states; and finally, it is hard to see how Virtual reality theory would work without extensive top-down signalling in the construction of the internal virtual reality model that can arise in the absence of bottom-up signals. We also note that studies of the global state of consciousness in anaesthesia suggest that a global state of consciousness depends on top-down signalling in the brain (Boly et al., 2011, 2012; Pal et al., 2020).

There are significant differences amongst the theories concerning how this top-down signalling should work, neurophysiologically and computationally. Some theories (e.g., Global neuronal workspace theory and versions of Higher-order thought and Metacognitive theories place the source of top-down signals more forward in the brain than others, such as Recurrent processing theory). In spite of these differences, the agreement on the need for top-down signalling is striking, and does seem to be supported by the available evidence.

Given that most theories of consciousness allocate key roles to uncertainty reduction and top-down signalling, and assuming these theories are at least roughly on the right track, it would be desirable if the systematic framework for the NCC was able to accommodate these two elements in a central but relatively theory neutral role. As we shall see, this requirement also makes PP a good fit.

5 Core properties of PP

Predictive processing is a theoretical framework that posits that the brain's overall function is to minimize long-term average prediction error. According to PP, the brain harbors a generative model of its internal, bodily environment as well as its external, worldly environment and continually finesses this model through a prediction/prediction error dynamic. The predictions are compared to the actual input, and the brain's job is to keep prediction error within expected bounds in the long term (Friston, 2010, 2005; Friston, Parr, et al., 2017). The underlying objective for this PP activity is for the organism to act (internally and externally) to maintain itself in its expected (homeostatic) states. As we shall acknowledge further in the final section, the PP framework is by no means universally accepted and there is substantial, interesting and challenging debate about it. Here, we seek to primarily develop its promising properties for the science of consciousness.

Predictive processing in the brain can occur in four basic ways, all analogous to familiar descriptions of scientific inference (for comprehensive reviews, see e.g., Bogacz, 2017; Buckley et al., 2017; Hohwy, 2020a; Yon et al., 2019). The central

principle for the operation of these four processes is that *on average and over the long run, (statistical) surprise* should be minimized:

Perception. The model can be revised in the light of prediction error, which will reduce the error over different timescales, at different levels of the model's hierarchy, and improve the model fit. On short timescales prediction errors are minimized to find the optimal (or near-optimal) posterior (using approximations to Bayesian inference). This is perceptual inference (Friston, 2003). On longer timescales the parameters of generative models can change, corresponding to perceptual learning, which in turn finesses models to perform better in perceptual inference. This is analogous to how a scientist might update their hypothesis in the light of new data, hoping to eventually learn models that generate predictions encountering only little prediction error.

Action. The model can be kept fixed and used to generate predictions even though prediction error may transiently increase. Action, or selective sampling of the sensory input, can then be used to minimize the prediction error. In particular, proprioceptive predictions are delivered to the classic reflex arcs, which engage the body plant until the expected sensory input is obtained. This is action, and is analogous to how a scientist may retain a hypothesis and control the environment for confounds until the expected evidence obtains (or, less optimistically, engages in repeating slight variations of an experiment until the desired result arrives). Action is then simply prediction error minimization with a different direction of fit than perceptual inference. In the PP framework, this is active inference and involves inference of policies for action, under expectations that precise (unambiguous) policies best minimize future prediction error (Friston et al., 2013; Friston, FitzGerald, Rigoli, Schwartenbeck, O'Doherty, et al., 2016). Conceiving of active inference in this way allows for both instrumental actions (actions that directly fulfil perceptual predictions) and exploratory, epistemic actions that improve the capacity for future instrumental action (Tschantz, Seth, et al., 2020). Importantly, actions can also be internal to the organism ('intero-actions'), so that interoceptive predictions enslave autonomic reflexes in order to maintain physiological homeostasis (Allen et al., 2019; Seth, 2013; Seth & Tsakiris, 2018). Active inference is indispensable to PP in any sufficiently rich understanding – a PP system without action will not be like any biological creature, certainly not like any system with a brain.

Attention. The dynamic exchange of predictions and prediction errors in PP can be modulated according to the expected precision of prediction error, relative to the prediction, such that prediction error minimization occurs on the basis of prediction error estimated to be trustworthy; this amounts to gain control (or gating), and functionally maps on to *attention* (Feldman & Friston, 2010; Hohwy, 2012). This is analogous to assessment of variance in statistical inference, and to how a scientist over time will build up expectations about which types of measurements will be precise and which not. Precision optimization refers to inference and learning for precision expectations. Note that despite the common association of

attention with exteroceptive modalities, precision optimization is ubiquitous in PP and applies (for example) equally to interoceptive inference (Petzschner et al., 2019).

Recognition, understanding, exploration. The model can be simplified (cf. complexity reduction) to enable prediction error minimization in the long run; this corresponds to Bayesian model selection, where complexity is penalized. It is analogous to how a scientist may prefer simpler models in the long run even though a more complex model may better fit the current evidence. Functionally, this may map on to aspects of recognition (rather than mere detection) of sensory causes and their interactions, within an understanding of the causal structure of the environment, where different models of what may have caused sensory input are weighed against each other. As mentioned above, in relation to epistemic action, an emphasis on minimizing future prediction error can lead naturally to notions of exploration, where prediction error is occasionally allowed to increase in order to enable new causal information to be discovered, opening new avenues for efficient future prediction error minimization (Friston, FitzGerald, Rigoli, Schwartenbeck, O’Doherty, et al., 2016; Friston et al., 2015; Tschantz, Seth, et al., 2020).

Putting these elements together, we arrive at the powerful idea that the brain embodies a model of its environment (including the organism itself), and exerts control on this environment through statistical inference geared towards minimizing its own long-term prediction error. This ambitious framework has been the focus of much work across neuroscience, cognitive science, biology and philosophy (e.g., Metzinger & Wiese, 2017; Hohwy, 2020a; Mendonça et al., 2020). The point of minimizing prediction error in the long run average, is that this allows the organism to stay within its expected states – i.e., to stay alive. This gives a perspective on the whole organism as primarily concerned with maintaining homeostasis by repeatedly visiting (expected) states of its (internal and external) environment. The overall prediction error minimization of the organism is therefore put into the context of the specific environmental impact (on physiological homeostasis) the organism is exposed to. This foundational biological perspective underwrites the claim that the PP framework is not in itself a theory of consciousness.

A key point about PP is that the mechanistic process of sensory prediction error minimization will (given certain assumptions such as normally distributed data and separable timescales) come to approximate Bayesian inference on the causes of sensory signals. This provides another perspective on PP: as transforming a typically intractable inference problem (exact Bayesian inference) into a tractable optimization problem. This perspective in turn licenses the view that a system that manages to minimize prediction error in the long run will tend to behave as if it were a representational system. This follows because Bayesian inference is representational, in the sense of resting on statistical models of the causes of data. (Already there is a link here to Chalmers’ example of a systematic NCC, mentioned earlier, as being a neural representational system, though PP is much richer than just this).

This perspective highlights that the PP framework can be cast at different levels of abstraction which make different claims about the underlying mechanism. On ‘strong’ interpretations, the brain is *in fact* utilizing generative models and approximate Bayesian inference to accomplish prediction error minimization. On ‘weaker’ interpretations, brains (and other systems) can be described as if they are utilizing such models and inferential processes, though the underlying physical-causal mechanism may be different (cf. the Watt governor; Van Gelder, 1995). Here lie interesting questions about the necessary and sufficient conditions required to *be* vs. to *have* a model, what this implies for PP, and for the application of PP as a framework for expressing systematic NCCs (Baltieri et al., 2020; Kiefer, 2017; Seth & Tsakiris, 2018; Van Es, 2020).

Note that for organisms living in a dynamically changing world, Bayesian inference will have to anticipate changes in the world that may threaten their prior beliefs (especially those prior beliefs concerning homeostasis). That is, such systems cannot just accumulate evidence under the assumption that the causes they are sampling from remain constant, they must take *volatility* into account too (i.e., changes in statistical distributions over time). This implies that Bayesian inference should be hierarchical: an adequate prediction-error minimizing system should find patterns in patterns and build corresponding expectations about expectations, be able to regulate the rate with which it learns, and it should be prepared to relinquish parts of the model that it expects to become obsolete (Mathys et al., 2011, 2014). In addition, with an appeal to hierarchical inference there is an opportunity to ameliorate the problem of understanding where priors come from: in a hierarchical setting, posteriors at one level form (empirical) priors for the level below. Notice that under PP, hierarchies should not be understood in overly idealized terms of neatly stacked levels passing messages only to their nearest neighbours. Cortical hierarchies, though mathematically describable, will recapitulate the causal structure of the world it is modelling, with several leap-frogging message pathways among more distant levels (Markov et al., 2013).

These core properties of the overall PP framework can be implemented mechanistically through *process theories* for particular kinds of systems, such as brains. The best known process theory for perceptual inference is predictive coding, and gradient descent on variational free energy for active inference (Friston, 2003; Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016). Predictive coding is a concrete ‘message passing’ scheme where neuronal representations in higher levels of cortical hierarchies generate predictions of representations in lower levels. The top-down predictions are compared with representations at the lower level to form a prediction error (associated with the activity of superficial pyramidal cells). This mismatch signal is passed back up the hierarchy, to update higher level representations (associated with the activity of deep pyramidal cells). As a result of this recursive exchange of signals, prediction errors are suppressed (‘explained away’) at all hierarchical levels, to provide a hierarchical explanation for sensory inputs at the very lowest (sensory) level. This message passing process encompasses a

limited number of signals, consistent with PP: top-down predictions and expected precisions, endogenous precision modulation, and bottom-up precision-weighted prediction error and precision prediction errors (canonical microcircuits for this scheme have been proposed Bastos et al., 2012; see also Walsh et al., 2020; and Friston, 2019b; see also Teufel & Fletcher, 2020 for possible roles for “bottom-up” signaling of predictions). Minimization of variational free energy for active inference is less widely known but is beginning to be extensively studied too (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016; Parr & Friston, 2018; Tschantz, Millidge, et al., 2020; Tschantz, Seth, et al., 2020). Outstanding issues in this area of research concern the underlying neuronal architecture of continuous vs. discrete processing, the relative extent to which predictive coding or complexity reduction occurs at higher or deeper levels of the cortical hierarchy (Friston, Parr, et al., 2017), how best to formalize notions of ‘expected free energy’ (Millidge et al., 2020), and – more generally – its relation to machine learning approaches such as reinforcement learning and ‘control as inference’ (Tschantz, Millidge, et al., 2020).

6 PP as a systematic NCC

Now we are in a position to propose PP as a promising candidate for satisfying the systematicity requirement for the NCC. Recall that to meet this goal we need a general framework for brain function with sufficient explanatory and predictive power to account for the structure and form of its correlated phenomenological distinctions. We further identified three challenges to the NCC search (global vs. local, screening off, and action) as well as two common elements of theories of consciousness (uncertainty reduction and top-down signal flow), which should be used as selection criteria for this general framework. PP evidently fits this bill. Here we show how this is the case assuming some ‘strong’ version of PP, that is, the existence of some explicit message passing scheme similar to that described for perceptual and active inference above. (This assumption is important because it opens the possibility that ‘weaker’ versions of PP will work less well to establish a systematic correlate for consciousness, in other words, PP will also be able to describe systems that are not conscious).

PP holds promise for addressing the ‘Global states and local contents’ challenge. PP is an account of the global state of creatures that act to remain in their expected states, and who in the course of this process install a probabilistic model of their internal and external environment. This perspective allows that local input, caused by the occurrence of particular environmental states, can be inserted into this global state, potentially leading to interactions between the global prediction error landscape and the local exposure to new causes. This perspective then connects specific contents to the global regulation of homeostatic and allostatic states (for discussion, see Stephan et al., 2016; Corcoran & Hohwy, 2018). This connection illustrates how the machinery of PP can be used to systematically account for properties of consciousness, without being a theory of consciousness

itself. From this perspective, global states of consciousness like sleep and anaesthesia may involve disruption of the top-down or bottom-up signaling that is needed to instantiate the process of prediction error minimization with respect to local exposure to sensory causes; sleep and dreaming may relate to how expected reductions in precision of sensory input lead to global reduced gain on prediction error (Hobson & Friston, 2012), and propofol-induced loss of consciousness seems to relate to disruption of top-down signaling (Boly et al., 2012).

PP holds promise also for the ‘Screening off: basic brain states and phenomenology’ challenge about how to ‘screen off’ less explanatory neural correlates from the more explanatory correlates. To address this challenge, we highlighted the importance of having ways of intervening on causal chains in neural systems that in a reasonable sense ‘carve nature at the joints’, that is, are informed by a mechanistic framework for brain function, and which are formulated at a level that can capture phenomenological distinctions.

PP meets the first part of this challenge since it posits a mechanism of prediction error minimization, replicated throughout levels of a cortical hierarchy. In other words, manipulations and interventions that are aimed at screening off for the purposes of the NCC, should be informed and interpreted by the process theories animating PP (we recognise that there remain many open theoretical and empirical questions about these process theories (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016; Heeger, 2017; Van der Helm, 2016)). Consider, as an example, an attempt to screen off upstream and downstream correlates for conscious face perception. In this example, it makes sense to ask how areas implicated in face perception, such as the fusiform face area (FFA) and the occipital face area (OFA) stand to each other in terms of hierarchical inference and message passing of predictions and prediction errors. It could be that even if activity in FFA is more tightly coupled to face perception than activity in OFA, we cannot understand the systematic contribution of FFA to conscious perception without appreciating the role of prediction error signals arriving from another area (such as OFA). In other words, it may be that PP can make contributions to the search for systematic NCCs by guiding the balancing of functional segregation (localization) and functional connectivity (integration) in formulating candidate NCCs. In a concrete instance of this kind of approach, Weilhhammer and colleagues (Weilhhammer et al., 2020) employed a predictive coding model of binocular rivalry to provide evidence that prefrontal activity is not downstream to alternations in conscious perception.

PP then also offers a powerful approach to the phenomenological part of this challenge, namely, to match neural substrates to the structure of conscious experience. Certain aspects of PP readily lend themselves to matching with informal descriptions of perceptual phenomenology. For example, conscious experiences typically contain many objects set in a contextualized scene. The hierarchical nature of PP nicely accommodates this phenomenological observation, so that sensory signals are ‘explained away’ by predictive models combining many degrees of invariance. This admits predictions about sensory causes that have different

degrees of invariance with respect to exogenous environmental dynamics and sensory changes resulting from actions. This is a useful property in light of the range of invariances characterizing perceptual phenomenology (e.g., objects are perceived invariantly as being a certain way while at the same time we experience their changing position, lighting, and even presence to our sensory surfaces at all during occlusion). PP also readily accounts for multisensory integration through approximation to precision-weighted optimal Bayesian inference (Alais & Burr, 2004; Hohwy, 2013: Ch. 5); it also speaks to unusual conscious experiences, for example in delusions and hallucinations, which may reflect sensory signals being inappropriately overwhelmed by perceptual priors (Corlett et al., 2019; Fletcher & Frith, 2009; Powers et al., 2017).

Perhaps most interestingly, the language of PP enables new questions (or reanimates old, neglected questions) about conscious phenomenology. For example, extending early claims that peripheral vision is ‘somehow statistical’ (Lettvin, 1976), we could ask whether perceptual content in the visual periphery is constituted by precision estimations, instead of merely being shaped by them. Or, starting from empirical observations rather than phenomenology, we could ask how the conceptual tools of PP account for intriguing findings such as the inability to distinguish between large variations in colour saturation for rapidly presented images (Cohen & Rubenstein, 2020). In other words, the language of PP enables new and productive mappings to be drawn between phenomenological descriptions and underlying neurocognitive processes. If PP can facilitate informative mapping of structured phenomenology with neural substrates screened off in an informed manner, then it can make significant headway on this challenge.

PP can also help address the ‘Action challenge’. This challenge asks for a flexible way to integrate action into an account of the structuring of conscious experience. PP does this first in a general way, by defining the organism in terms of how it acts to maintain itself in its expected states. Within this general setting, PP can account for the specific role of action in perception through the notion of active inference where policies are inferred on the basis of the expected prediction error, given the selected policy. Policies cover long-term goals such as going shopping, or choosing an education (or maintaining homeostatic and interoceptive states in the face of environmentally or internally caused fluctuations), and translate these goals through shorter-term policies all the way down to control states that can be executed by the agent’s body (such as limb or eye movement). The key move here is to recognize that goals (in a reinforcement learning sense) can be expressed as priors within PP, so that an organism will come to act in ways that fulfil its (biased) expectations about its future condition (Friston, FitzGerald, Rigoli, Schwartenbeck, O’Doherty, et al., 2016; Tschantz, Millidge, et al., 2020).

The fact that both perception and action are inferential, albeit with different directions of fit, places them on equal footing and ensures they can meaningfully participate within the same, systematic NCC project. This can be exemplified in a computational model that explains one of the basic paradigms in consciousness

science, binocular rivalry, in terms of active inference where selective sampling, driven by beliefs about volatility, leads to perceptual alternations (Parr et al., 2019). In addition, since, within PP, actions themselves emerge from fulfillment of (proprioceptive) predictions, the PP framework provides a means of accounting for the phenomenology of action – such as sense of agency; (Pacherie, 2008) as well as the influence of action on other conscious contents (such as the fidgeting we experience when bored or excited (Perrykkad & Hohwy, 2020), or the role of action-related predictions in the shaping of conscious access (Skora et al., 2020). PP is therefore well placed to systematically integrate an action component into the NCC, and thereby to afford an understanding of the key role for action in consciousness, rather than conceiving action as an optional extra to consciousness.

In addition to the three challenges for a systematic NCC, we also identified two constraints, extracted from theories of consciousness, concerning uncertainty reduction and top-down signaling.

PP can easily provide a systematic framework for the NCC that identifies uncertainty reduction as a core element. The reason is simple: prediction error minimization, construed as (implicit) minimization of surprisal, is equivalent to uncertainty reduction over the long-term average. Prediction error minimization happens over multiple hierarchical, spatiotemporal scales, and through several interlinked processes (perceptual and active inference, precision optimization, complexity reduction); this richness affords some capacity to capture the different styles of uncertainty reduction that various theories of consciousness focus on. Thus PP seems consistent with the following ideas: that (phenomenal) consciousness correlates with recurrent processing within perceptual cortices (as opposed to requiring the contribution of domain-general, higher-order, or ‘executive’ cortices) (cf. Recurrent processing theory); that (access) consciousness requires ignition into a more anterior global workspace, perhaps once a prediction error minimization threshold has been crossed (cf. Global neuronal workspace theory; this leaves an interesting open question for PP as to what determines this threshold, with one option being that it varies contextually with expected precision); that a conscious experience is the way it is, and not some other way, in part because of the repertoire of excluded alternatives (cf. Integrated information theory; though much formal work still needs to be done to ascertain in which ways uncertainty reduction in PP and integrated information theory may align); or that metacognitive assessment of first-order representations requires prefrontal areas (cf. Higher order thought and Metacognitive theories). Note, we are not claiming that all these theories reduce to PP. Instead, we argue that several of them are consistent with different aspects of PP (at least in their theoretical motivation, if not in formal detail), and – further – that PP’s tools may be useful in broadening the conceptual appeal and empirical grip of some of these theories.

PP is also well placed to underpin the second constraint: top-down signalling. This is because most prediction error minimization schemes, such as hierarchical predictive coding, are cashed out in ways that emphasise top-down message pass-

ing. It may be that in some (e.g., single celled) PP organisms, there is no straightforward way to distinguish anything resembling top-down or bottom-up signalling. But it is reasonably clear that directionally-oriented message passing does occur in primates and many other animals, and the prevalence of top-down or ‘inside out’ connections in the brain is indeed one of the main motivations for PP in the first place (e.g., Friston, 2003). While this basic functional architecture is undeniably prominent, many open questions remain, the resolution of which likely deliver a more complicated picture. For example, it may be that different message passing schemes are required for continuous and discrete processing (Friston, Parr, et al., 2017), and there may be a role for ‘bottom-up predictions’ (Teufel & Fletcher, 2020), perhaps in relation to ‘amortized’ inference in machine learning. From the perspective of developing a systematic NCC, these questions should be considered in relation to their relevance for phenomenological properties of consciousness.

In short, we find that PP is an exceptionally promising candidate to satisfy the systematicity requirement in the search for candidate NCCs. It is a general-purpose mechanistic framework, which has plausible neurobiological underpinnings, and which is well equipped to address significant challenges facing the NCC project. Moreover, it is able to marry the NCC concept to key considerations in some prominent theories of consciousness, and its constructs already show potential in accounting for fundamental aspects of conscious phenomenology. By interpreting NCC findings in terms of something that is not itself, in the first instance, a theory of consciousness, it therefore seems possible to increase the predictive and explanatory value of the NCC.

Of course, something that has great predictive and explanatory power with respect to x may itself be, or become, a theory of x . And we view it as indeed possible (even likely) that, in time, some aspects of PP will emerge as a theoretical explanation of consciousness. Thus, PP could eventually supersede current theories of consciousness. For this to happen, a core challenge will be to distinguish those aspects of PP that apply particularly to conscious creatures, from those that don’t. In its broadest and most general formulation, in terms of the free energy principle (Friston, 2019a), the theory applies to any system in a far-from-equilibrium steady state, which almost certainly includes non-conscious creatures – indeed it may apply literally to every *thing*. The strategy should therefore be to learn what is special about the way in which prediction error is minimized in creatures that we are relatively sure are conscious, such as primates (probably all mammals), and with this understanding to formulate a theory of at least some aspects of consciousness in terms of PP (Wiese, 2018, 2020). There are various possible avenues here, but to give this prospect some substance, it may be that conscious selfhood is grounded in control-oriented active inference of physiological homeostasis – a “beast-machine” theory of consciousness (Seth, 2021; Seth & Tsakiris, 2018). We return to this theme in the concluding remarks below.

7 Predictive and explanatory power of PP as a systematic basis for NCC

Implementation of PP to arrive at a systematic NCC can happen on several fronts, with much relevant work already underway. We give a flavor of some of this work below.

At the theoretical level, there are emerging basic formulations of predictive processing (or active inference, or the free energy principle, or adjacent or kindred formulations) directed at consciousness (or sentience, or self) (Carhart-Harris et al., 2014; Chang et al., 2020; Clark, 2019; Friston, 2018; Friston, Wiese, et al., 2020; Rudrauf et al., 2017; Williford et al., 2018), including the emergence of the problem of consciousness itself (Clark et al., 2019), or how PP might supersede, dissolve, or deflate the ‘hard problem’ of consciousness (Clark, 2019; Seth, 2016, 2021). There has been a significant amount of research mapping a large set of properties of PP to many different aspects of conscious phenomenology (Clark, 2015; Hohwy, 2013, 2020a; Seth, 2019). Existing theories of consciousness are being interpreted in the light of PP, such as Heterophenomenology (Dołęga & Dewhurst, 2020), Global neuronal workspace theory (Hohwy, 2013; Whyte, 2019; Whyte & Smith, 2020), versions of Higher-order thought theory and Metacognitive theories (Hohwy, 2015; Sandved Smith et al., 2020), Attention-based theories (Marchi & Hohwy, 2020), and Integrated information theory (for related discussion, see Kanai et al., 2019).

Moving closer to empirical implementation, PP is increasingly able to speak to key paradigms in consciousness science, such as binocular rivalry. An early contribution here proposed that if some representational content enters consciousness, then this content is captured in the perceptual hypothesis that is best at minimizing prediction error; or, equivalently, for content to be conscious it must be represented by the posterior with the most evidence (highest probability) (Hohwy et al., 2008). Note that this is only a starting point since it does not pretend to be a sufficient condition for consciousness (and even with that caveat, there are some additional questions about how active inference may contribute to this idea, as the action challenge highlights; for example, desired or survival-relevant states might influence the ‘winning’ posterior through increased prior precision (Seth & Tsakiris, 2018)). The idea that the best perceptual hypothesis determines conscious content has been pursued successfully with quantitative predictions and neuroimaging methods (Weilnhammer et al., 2017), pointing to involvement of prefrontal cortex in the lead-up to perceptual transitions (Weilnhammer et al., 2020). This position has since been re-visited in the computational model mentioned earlier, which integrates active inference and thereby displays the potential for PP to see conscious perception in its essential relation to action (Parr et al., 2019).

One aspect of PP that applies across many empirical contexts is the association of top-down signaling with predictions (or expectations) that enable and shape perceptual content. This sharply defined functional role contrasts with some other perspectives, like Recurrent processing theory, in which the role of top-down sig-

naling is left comparatively unspecified. Following early demonstrations that the integrity of top-down signaling within visual cortex was necessary for conscious perception of motion (Pascual-Leone & Walsh, 2001), there has been a productive upsurge in studies of how expectations shape perception, much of which has been directly inspired by versions of PP (Lange et al., 2018; Summerfield & Egner, 2009). Behavioural experiments have addressed the simple question of whether we (are more likely to) consciously see what we expect to see, or what violates our expectations. Pinto and colleagues found evidence using continuous flash suppression that valid perceptual expectations accelerated conscious access (Pinto et al., 2015); Melloni and colleagues used a perceptual hysteresis design to show that valid expectations facilitated conscious perception and altered the latency of event-related potentials distinguishing seen from unseen stimuli (Melloni et al., 2011); this focus extends to action, where Yon and colleagues found that action sharpens the sensory representations of expected outcomes (Yon et al., 2018), and where Suzuki and colleagues found that valid sensorimotor predictions accelerated breakthrough of virtual objects from continuous flash suppression (Suzuki et al., 2019; see also Skora et al., 2020).

A growing number of studies using fMRI have provided evidence for how contextual expectations modulate activity in a content specific manner in early visual areas even in the absence of sensory input to those areas, and have begun unraveling how the interplay of predictions and prediction errors might take place across distinct laminae of the cortex (Kok et al., 2019; Muckli et al., 2015). In one recent study, expectations based on word context were found to both increase the perceptual legibility of individual letters, and to sharpen their representations within early visual cortex (Heilbron et al., 2020). Another study, using MEG, was able to quantify the time course of the influence of probabilistic top-down priors in perception (Aru et al., 2016; Dijkstra et al., 2020). A further study systematically explored the dynamic interactions of predictions for conscious awareness (Meijs et al., 2018). A number of studies have applied PP to the question of aberrant conscious perception, for example, in schizophrenia (Cole et al., 2020; Stuke et al., 2019) and autism (Lawson et al., 2017; Skewes et al., 2015).

An important question in consciousness science is the relation between consciousness and attention. Theories of the Attention-based type suggest a tight relation (Graziano et al., 2020; Prinz, 2012), while some theoretical and empirical considerations suggest that consciousness and (at least top-down) attention can be dissociated (Koch & Tsuchiya, 2007). PP seems able to provide a middle ground allowing a tight connection between attention and consciousness together with elements of dissociation, via the multifaceted involvement of precision optimisation (Hohwy, 2012). As mentioned, PP sees attention as precision optimisation, which is required for any perceptual and active inference in a changing world. This means that any processing leading to any conscious perception must include attention, at least in this PP-sense, to some degree.

Methodologically, PP has inspired studies of attention and expectation treated as orthogonal factors. For example, one influential study used PP to derive hy-

potheses about interactions between attention and expectation, and explored how these interactions would manifest for conscious perception at low visual processing areas of the brain (V1); (Kok et al., 2012). Dual-task paradigms can also be used to dissociate the effects of expectation from those of attention (e.g., Sherman et al., 2015, in a study connecting prior expectations to metacognition). Recent studies using frequency tagging in EEG have sought to use PP to disentangle the top-down and bottom-up pathways for expectation and attention (Gordon et al., 2019).

There is increasing focus on Metacognitive approaches to consciousness. There is a clear sense in which PP involves a certain kind of metacognition, namely in terms of higher-level expected precisions. That is, the PP system is constantly 'looking down' at its own representations seeking to learn how their precisions (and more generally their sufficient statistics) change over time. Many questions are still open about how this general PP element can speak to the role of different kinds of metacognition in consciousness (for deployment of active inference to metacognition, see Sandved Smith et al., 2020). An outstanding challenge here will be to formally integrate the framework of signal detection theory, in which most theories of metacognition are articulated (e.g., Barrett et al., 2013) with PP (see, e.g., FitzGerald et al., 2014; Fleming, 2020).

As we look briefly to more subtle aspects of conscious experience, PP is surprisingly resourceful. Conscious perception combines both invariant and variant elements, and is mostly accompanied by a sense of presence of the perceived objects and properties; this may be accounted for by the way states of affairs are represented throughout the cortical hierarchy, which exhibits a gradient from variance to invariance (Clark, 2015; Felleman & Essen, 1991; Hohwy, 2013), which appears to underwrite perceptual stability in the visual cortex (Podvalny et al., 2017). A further contributor to presence could be the ability, in some PP systems, for counterfactual prediction error minimization (Seth, 2014) allowing representational access to occluded properties. Since PP applies equally to body-related perceptual inference (e.g., inference on the causes of proprioceptive and interoceptive signals), it is able to address issues about bodily self-awareness (Allen et al., 2019; Seth, 2013). The role of action for self-representation is also being explored with PP; in particular, sensory attenuation (manifested, e.g., in our inability to tickle ourselves) in action has received a new explanation in terms of precision optimization (Brown et al., 2013; Van Doorn et al., 2015; Vasser et al., 2019).

This overall approach to self-related inference may be extended to models of the philosophically recalcitrant topic of the self (Apps & Tsakiris, 2014; Hohwy & Michael, 2017; Limanowski & Blankenburg, 2013); two recent collections have several relevant PP contributions on this topic (Milliere & Metzinger, 2020; Tsakiris & Preester, 2018). Indeed, ambitious applications of PP have begun to account for the subjective experiences of embodied selfhood as arising from active inference of interoceptive states, where such inference is motivated by the control or regulation of the hidden causes of these states rather than by their accurate representation (Seth, 2015b, 2015a; Wiese, 2015). These ideas call on early cybernetic formulations

of predictive regulation of essential variables, like those of Ross Ashby (Ashby, 1954), which provide an alternative context for the free energy principle.

It is worth sounding a note of caution about interpreting evidence for (especially) altered experiences of embodiment in terms of Bayesian inference on body-related sensory signals. For example, a recent study by Lush and colleagues has revealed a link between subjective reports of illusory ownership in the well-known ‘rubber hand illusion’ (RHI) and hypnotic suggestion (Lush et al., 2020). In this study, individual differences in the strength of the RHI correlated tightly with individual differences in hypnotic suggestibility (which the authors refer to as trait ‘phenomenological control’). This result raises the possibility that reported experiences of embodiment of the RHI may be due, at least in part, to implicit imaginative suggestions arising from the experimental context, putting pressure on accounts of the RHI that are based primarily on multisensory integration. However, the PP view offers a potential solution to this confound too, by considering suggestion effects as top-down expectations of a distinctive kind (see, for example, Martin & Pacherie, 2019).

As a final example, the sense of the present as temporally extended may be explained by PP-like internal simulation (Grush, 2006); the sense of temporal flow of conscious experience – how one moment passes seamlessly into the next even when nothing much happens in the environment – may be explained by the dynamics of hierarchical Bayes that encode expected change in the environment (Hohwy et al., 2016; Sherman et al., 2020; Vogel et al., 2020, 2019); this process may in turn relate to the alternation of percepts in binocular rivalry, as discussed above.

In summary, then, PP is already able to cash out significant predictive and explanatory value for the study of consciousness. This supports the proposal that PP is an excellent candidate for providing the foundational framework for a systematic NCC. It also encourages the prospect that, eventually, aspects of PP may themselves coalesce into a theory of consciousness in its own right.

8 Challenges and concluding remarks

We have proposed that the science of consciousness can move toward a more explanatory and predictive conception of the NCC by capitalising on the predictive processing (PP) framework. PP holds unique promise for addressing challenges that a successful NCC strategy must overcome, and for linking candidate NCCs to key theoretical constraints shared among several prominent theories of consciousness. We are not aware of any other theoretical framework with this kind of potential for moving toward a systematic NCC.

Several areas of theoretical and empirical research are already demonstrating the potential of PP as such a basis for a systematic, rather than arbitrary, NCC. Full exploitation of this potential will be facilitated by focused efforts to draw on the PP framework when formulating and testing for candidate NCCs. Altogether, this approach will deliver significantly enhanced predictive and explanatory value

for the NCC program, as well as for the science of consciousness at large. In the end, this approach may also change the extent to which the hard problem of consciousness is still seen, by some, as a fundamental limit on a materialist science of consciousness.

In this discussion we have focused on the promising explanatory and unificatory aspects of PP. However, PP, although increasingly pervasive, is certainly not universally accepted and there are conflicting findings and perspectives, driving ongoing debate (e.g., Alilović et al., 2019; Cao, 2020; Li & Ma, 2020; Litwin & Milkowski, 2020; Rahnev & Denison, 2018). For example, one recurring challenge to PP is that worry that an imperative to minimize sensory prediction errors will lead to agents that do nothing at all – the so-called ‘dark room problem’ (Friston et al., 2012; Sun & Firestone, 2020a, 2020b). However, this challenge has already been effectively addressed by extensions of PP that consider the minimization of expected future prediction error, rather than just prediction error in the here and now (Hohwy, 2020b; Seth et al., 2020; Van de Cruys et al., 2020). Overall, PP as a framework continues to evolve, with new formal and modelling approaches appearing at a steady pace, some of which are yet to find its way into the broader debates about PP (Friston, Parr, et al., 2017; Friston, Da Costa, et al., 2020).

These new approaches will plausibly help generate robust experimental results, and could also enable closer links between neural substrates and conscious phenomenology. We believe the incorporation of active inference into consciousness science will be one major step in this revision, and that an increased focus on interoceptive inference will bring us closer to conscious phenomenology, especially in relation to affect, emotion, and self (Seth, 2021). Likewise, it will be important to gain clarity on the distinction between continuous and discrete predictive processing, both in order to reveal neural substrates and message passing and to begin exploring the phenomenology of thinking and (propositional) reasoning.

A foundational challenge to PP is the question whether the brain is actually engaged in approximate Bayesian inference, or whether this is just an attractive framework within which to interpret behavior and brain dynamics. This relates to the question whether organisms have explicit internal models, implementing PP processes, or whether their body plant in a more abstract sense embodies an implicit model. The answers to these questions will help determine how PP can underwrite the search for NCCs, and in what sense it will be of explanatory and predictive value to do so. These answers might also afford an opportunity to begin addressing a related challenge, namely why and how PP should underpin consciousness at all, if it is so general that it literally underpins every organism’s existence – including those that are not conscious (e.g., viruses)? A first approach to this challenge would be to acknowledge that even if PP (in its free energy formulation) is indeed of maximal scope, the process theories under PP apply to particular kinds of systems and thereby allow for differences in the way PP is implemented. Such differences could in principle speak to the distinction between conscious and non-conscious creatures. In addition to the distinction between explicit and im-

PLICIT generative models, a further useful distinction could be between creatures with and without a capacity for explicit counterfactual modelling of expected prediction error, or with and without temporally deep generative models (Friston, 2017). Such differences may potentially also account for differences in global conscious states in conscious creatures – such as the difference between anaesthesia and normal wakefulness in humans.

It remains to be seen if developments in these respects will help address important clinical and ethical issues for individuals and creatures at the edges of consciousness, such as disorders of consciousness, infants, and non-verbal animals. It also remains to be seen if the accumulation of these PP-driven advances in consciousness science will eventually lead to meaningful inroads on the hard problem of consciousness, or onto the meta-problem of why many people are convinced that there is a hard problem of consciousness (Chalmers, 2018). It is precisely because PP is not itself a theory of consciousness that it holds unique promise for shedding new light on these perennially intractable issues.

Acknowledgments

AKS is grateful to the Dr. Mortimer and Theresa Sackler Foundation, which supports the Sackler Centre for Consciousness, and to the CIFAR Program on Brain, Mind, and Consciousness. JH and AKS are supported by the Australian Research Council Discovery grant DP160102770.

References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14(3), 257–262. <https://doi.org/10.1016/j.cub.2004.01.029>
- Alilović, J., Timmermans, B., Reteig, L. C., van Gaal, S., & Slagter, H. A. (2019). No evidence that predictions and attention modulate the first feedforward sweep of cortical information processing. *Cerebral Cortex*, 29(5), 2261–2278. <https://doi.org/10.1093/cercor/bhz038>
- Allen, M., Levy, A., Parr, T., & Friston, K. J. (2019). In the body’s eye: The computational anatomy of interoceptive inference. *bioRxiv*, 603928. <https://doi.org/10.1101/603928>
- Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, 41, 85–97. <https://doi.org/10.1016/j.neubiorev.2013.01.029>
- 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>
- Aru, J., Rutiku, R., Wibral, M., Singer, W., & Melloni, L. (2016). Early effects of previous experience on conscious perception. *Neuroscience of Consciousness*, 2016(1). <https://doi.org/10.1093/nc/niw004>
- Aru, J., Suzuki, M., Rutiku, R., Larkum, M. E., & Bachmann, T. (2019). Coupling the state and contents of consciousness. *Frontiers in Systems Neuroscience*, 13. <https://doi.org/10.3389/fnsys.2019.00043>
- Ashby, W. R. (1954). *Design for a brain*. Wiley.
- Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.
- Bachmann, T. (2012). How to begin to overcome the ambiguity present in differentiation between contents and levels of consciousness? *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00082>
- Baltieri, M., Buckley, C. L., & Bruineberg, J. (2020). Predictions in the eye of the beholder: An active inference account of Watt governors. *Artificial Life Conference Proceedings*, 32, 121–129. https://doi.org/10.1162/isal_a_00288
- Barrett, A. B., Dienes, Z., & Seth, A. K. (2013). Measures of metacognition on signal-detection theoretic models. *Psychological Methods*, 18(4), 535–552. <https://doi.org/10.1037/a0033268>

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, 1(II), 3. <https://doi.org/10.33735/phimisci.2020.II.64>



- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711. <https://doi.org/10.1016/j.neuron.2012.10.038>
- 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., Cleeremans, A., & Wilken, P. (Eds.). (2009). *Oxford companion to consciousness*. Oxford University Press.
- Bayne, T., Seth, A. K., & Massimini, M. (2020). Are there islands of awareness? *Trends in Neurosciences*, *43*(1), 6–16. <https://doi.org/10.1016/j.tins.2019.11.003>
- Bechtel, W. (2007). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Taylor & Francis.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247. <https://doi.org/10.1017/S0140525X00038188>
- Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, *76*, 198–211. <https://doi.org/10.1016/j.jmp.2015.11.003>
- Boly, M., Garrido, M. I., Gosseries, O., Bruno, M.-A., Boveroux, P., Schnakers, C., Massimini, M., Litvak, V., Laureys, S., & Friston, K. J. (2011). Preserved feedforward but impaired top-down processes in the vegetative state. *Science*, *332*(6031), 858–862. <https://doi.org/10.1126/science.1202043>
- Boly, M., Moran, R., Murphy, M., Boveroux, P., Bruno, M.-A., Noirhomme, Q., Ledoux, D., Bonhomme, V., Brichant, J.-F., Tononi, G., Laureys, S., & Friston, K. J. (2012). Connectivity changes underlying spectral EEG changes during propofol-induced loss of consciousness. *Journal of Neuroscience*, *32*(20), 7082–7090. <https://doi.org/10.1523/JNEUROSCI.3769-11.2012>
- Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K. J. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, *14*(4), 411–427. <https://doi.org/10.1007/s10339-013-0571-3>
- Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences*, *23*(9), 754–768. <https://doi.org/10.1016/j.tics.2019.06.009>
- Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, *81*, 55–79. <https://doi.org/10.1016/j.jmp.2017.09.004>
- Cao, R. (2020). New labels for old ideas: Predictive processing and the interpretation of neural signals. *Review of Philosophy and Psychology*, *11*(3), 517–546. <https://doi.org/10.1007/s13164-020-00481-x>
- Carhart-Harris, R. L., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., Chialvo, D. R., & Nutt, D. (2014). The entropic brain: A theory of conscious states informed by neuroimaging research with psychedelic drugs. *Frontiers in Human Neuroscience*, *8*. <https://doi.org/10.3389/fnhum.2014.00020>
- Chalmers, D. J. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.), *Neural correlates of consciousness* (pp. 17–39). MIT Press.
- Chalmers, D. J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, *25*(9-10), 6–61.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Chang, A. Y. C., Biehl, M., Yu, Y., & Kanai, R. (2020). Information closure theory of consciousness. *Frontiers in Psychology*, *11*. <https://doi.org/10.3389/fpsyg.2020.01504>
- Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.
- Clark, A. (2019). Consciousness as generative entanglement. *Journal of Philosophy*, *116*(12), 645–662. <https://doi.org/10.5840/jphil20191161241>
- Clark, A., Friston, K. J., & Wilkinson, S. (2019). Bayesing qualia: Consciousness as inference, not raw datum. *Journal of Consciousness Studies*, *26*(9-10), 19–33.
- Cleeremans, A. (2011). The radical plasticity thesis: How the brain learns to be conscious. *Frontiers in Psychology*, *2*. <https://doi.org/10.3389/fpsyg.2011.00086>
- Cohen, M. A., & Rubenstein, J. (2020). How much color do we see in the blink of an eye? *Cognition*, *200*, 104268. <https://doi.org/10.1016/j.cognition.2020.104268>
- Cole, D. M., Diaconescu, A. O., Pfeiffer, U. J., Brodersen, K. H., Mathys, C. D., Julkowsky, D., Ruhrmann, S., Schilbach, L., Tittgemeyer, M., Vogeley, K., & Stephan, K. E. (2020). Atypical processing of uncertainty in individuals at risk for psychosis. *NeuroImage: Clinical*, *26*, 102239. <https://doi.org/10.1016/j.nicl.2020.102239>
- Corcoran, A. W., & Hohwy, J. (2018). Allostasis, interoception, and the free energy principle: Feeling our way forward. In M. Tsakiris & H. de Preester (Eds.), *The interoceptive basis of the mind* (pp. 272–292). Oxford University Press.
- Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., & Powers, A. R. (2019). Hallucinations and strong priors. *Trends in Cognitive Sciences*, *23*(2), 114–127. <https://doi.org/10.1016/j.tics.2018.12.001>
- Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press.

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, *1*(II), 3. <https://doi.org/10.33735/phimisci.2020.II.64>



- Crick, F., & Koch, C. (1990a). Some reflections on visual awareness. *Sympos Quant Biol*, 953–962. <https://doi.org/10.1101/sqb.1990.055.01.089>
- Crick, F., & Koch, C. (1990b). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2, 263–275.
- Damasio, A. (2000). *The feeling of what happens: Body and emotion in the making of consciousness*. Mariner Books.
- De Haan, E. H. F., Corballis, P. M., Hillyard, S. A., Marzi, C. A., Seth, A., Lamme, V. A. F., Volz, L., Fabri, M., Schechter, E., Bayne, T., Corballis, M., & Pinto, Y. (2020). Split-brain: What we know now and why this is important for understanding consciousness. *Neuropsychology Review*, 30(2), 224–233. <https://doi.org/10.1007/s11065-020-09439-3>
- Dehaene, S. (2011). Conscious and nonconscious processes: Distinct forms of evidence accumulation? *Biological Physics. Progress in Mathematical Physics*, 60. https://doi.org/10.1007/978-3-0346-0428-4_7
- Dennett, D. C. (1991). *Consciousness explained*. Little, Brown & Company.
- Descartes, R. (1641). *Descartes: Meditations on first philosophy: With selections from the objections and replies* (J. Cottingham, Ed.). Cambridge University Press.
- Dijkstra, N., Ambrogioni, L., Vidaurre, D., & Gerven, M. van. (2020). Neural dynamics of perceptual inference and its reversal during imagery. *eLife*, 9, e53588. <https://doi.org/10.7554/eLife.53588>
- Doerig, A., Schurger, A., & Herzog, M. H. (2020). Hard criteria for empirical theories of consciousness. *Cognitive Neuroscience*, 1–22. <https://doi.org/10.1080/17588928.2020.1772214>
- Dołęga, K., & Dewhurst, J. E. (2020). Fame in the predictive brain: A deflationary approach to explaining consciousness in the prediction error minimization framework. *Synthese*. <https://doi.org/10.1007/s11229-020-02548-9>
- Engel, A. K., Friston, K. J., & Kragic, D. (Eds.). (2016). *The pragmatic turn: Toward action-oriented views in cognitive science*. The MIT Press.
- Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, 5(1), 16–25. [https://doi.org/10.1016/S1364-6613\(00\)01568-0](https://doi.org/10.1016/S1364-6613(00)01568-0)
- Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4. <https://doi.org/10.3389/fnhum.2010.00215>
- Felleman, D. J., & Essen, D. C. V. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, 1–47. <https://doi.org/10.1093/cercor/1.1.1>
- Fernández-Espejo, D., & Owen, A. M. (2013). Detecting awareness after severe brain injury. *Nature Reviews Neuroscience*, 14(11), 801–809. <https://doi.org/10.1038/nrn3608>
- Fink, S. B. (2016). A deeper look at the “neural correlate of consciousness”. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01044>
- FitzGerald, T. H. B., Schwartenbeck, P., Moutoussis, M., Dolan, R. J., & Friston, K. J. (2014). Active inference, evidence accumulation, and the urn task. *Neural Computation*, 27(2), 306–328. https://doi.org/10.1162/NECO_a_00699
- Fleming, S. M. (2020). Awareness as inference in a higher-order state space. *Neuroscience of Consciousness*, 2020(1). <https://doi.org/10.1093/nc/niz020>
- Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-specific disruption of perceptual confidence. *Psychological Science*. <https://doi.org/10.1177/0956797614557697>
- Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10(1), 48–58. <https://doi.org/10.1038/nrn2536>
- Friston, K. J. (2017). The mathematics of mind-time. *Aeon*. <https://aeon.co/essays/consciousness-is-not-a-thing-but-a-process-of-inference>
- Friston, K. J. (2018). Am I self-conscious? (Or does self-organization entail self-consciousness?). *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00579>
- Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>
- Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. J. (2019a). A free energy principle for a particular physics. *arXiv*. <https://doi.org/10.1101/201906.10184>
- Friston, K. J. (2019b). Waves of prediction. *PLoS Biology*, 17(10), e3000426. <https://doi.org/10.1371/journal.pbio.3000426>
- Friston, K. J. (2003). Learning and inference in the brain. *Neural Networks*, 16(9), 1325–1352. <https://doi.org/10.1016/j.neunet.2003.06.005>
- Friston, K. J., Da Costa, L., Hafner, D., Hesp, C., & Parr, T. (2020). Sophisticated inference. *arXiv*. <https://doi.org/10.1101/202006.04120>
- Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O’Doherty, J., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, 68, 862–879. <https://doi.org/10.1016/j.neubiorev.2016.06.022>

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, 1(II), 3. <https://doi.org/10.33735/phimisci.2020.II.64>



- Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016). Active inference: A process theory. *Neural Computation*, 29(1), 1–49. https://doi.org/10.1162/NECO_a_00912
- Friston, K. J., Parr, T., & Vries, B. de. (2017). The graphical brain: Belief propagation and active inference. *Network Neuroscience*, 1(4), 381–414. https://doi.org/10.1162/NETN_a_00018
- Friston, K. J., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, 6(4), 187–214. <https://doi.org/10.1080/17588928.2015.1020053>
- Friston, K. J., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: Active inference and agency. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00598>
- Friston, K. J., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00130>
- Friston, K. J., Wiese, W., & Hobson, J. A. (2020). Sentience and the origins of consciousness: From Cartesian duality to Markovian monism. *Entropy*, 22(5), 516. <https://doi.org/10.3390/e22050516>
- Gold, I. (1999). Does 40-Hz oscillation play a role in visual consciousness? *Consciousness and Cognition*, 8(2), 186–195. <https://doi.org/10.1006/ccog.1999.0399>
- Gordon, N., Tsuchiya, N., Koenig-Robert, R., & Hohwy, J. (2019). Expectation and attention increase the integration of top-down and bottom-up signals in perception through different pathways. *PLOS Biology*, 17(4), e3000233. <https://doi.org/10.1371/journal.pbio.3000233>
- Graaf, T. A. de, 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>
- Graziano, M. S. A., Guterstam, A., Bio, B. J., & Wilterson, A. I. (2020). Toward a standard model of consciousness: Reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cognitive Neuropsychology*, 37(3-4), 155–172. <https://doi.org/10.1080/02643294.2019.1670630>
- Graziano, M. S. A., & Webb, T. W. (2015). The attention schema theory: A mechanistic account of subjective awareness. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00500>
- Grush, R. (2006). How to, and how not to, bridge computational cognitive neuroscience and Husserlian phenomenology of time consciousness. *Synthese*, 153(3), 417–450. <https://doi.org/10.1007/s11229-006-9100-6>
- Hameroff, S., & Penrose, R. (2014). Consciousness in the universe: A review of the “Orch OR” theory. *Physics of Life Reviews*, 11(1), 39–78. <https://doi.org/10.1016/j.plrev.2013.08.002>
- Haun, A., & Tononi, G. (2019). Why does space feel the way it does? Towards a principled account of spatial experience. *Entropy*, 21(12), 1160. <https://doi.org/10.3390/e21121160>
- Heeger, D. J. (2017). Theory of cortical function. *Proceedings of the National Academy of Sciences*, 114(8), 1773–1782. <https://doi.org/10.1073/pnas.1619788114>
- Heilbron, M., Richter, D., Ekman, M., Hagoort, P., & Lange, F. P. de. (2020). Word contexts enhance the neural representation of individual letters in early visual cortex. *Nature Communications*, 11(1), 321. <https://doi.org/10.1038/s41467-019-13996-4>
- Hobson, J. A., & Friston, K. J. (2012). Waking and dreaming consciousness: Neurobiological and functional considerations. *Progress in Neurobiology*, 98(1), 82–98. <https://doi.org/10.1016/j.pneurobio.2012.05.003>
- Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00096>
- Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
- Hohwy, J. (2015). Prediction error minimization, mental and developmental disorder, and statistical theories of consciousness. In *Disturbed consciousness: New essays on psychopathology and theories of consciousness* (pp. 293–324). MIT Press.
- Hohwy, J. (2020a). New directions in predictive processing. *Mind & Language*, 35(2), 209–223. <https://doi.org/10.1111/mila.12281>
- Hohwy, J. (2020b). Self-supervision, normativity and the free energy principle. *Synthese*. <https://doi.org/10.1007/s11229-020-02622-2>
- 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.
- Hohwy, J., & Michael, J. (2017). Why does any body have a self? In F. de Vignemont & A. J. T. Alsmith (Eds.), *The body and the self, revisited* (pp. 363–391). MIT Press.
- Hohwy, J., Paton, B., & Palmer, C. (2016). Distrusting the present. *Phenomenology and the Cognitive Sciences*, 15(3), 315–335. <https://doi.org/10.1007/s11097-015-9439-6>

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, 1(II), 3. <https://doi.org/10.33735/phimisci.2020.II.64>



- Hohwy, J., Roepstorff, A., & Friston, K. J. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, 108(3), 687–701. <https://doi.org/10.1016/j.cognition.2008.05.010>
- Hurley, S. L. (1998). *Consciousness in action*. Harvard University Press.
- Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. Cambridge, Mass.: MIT Press.
- Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly* 32, 32(127), 127–136. <https://doi.org/10.2307/2960077>
- Kanai, R., Chang, A., Yu, Y., Magrans de Abril, I., Biehl, M., & Guttenberg, N. (2019). Information generation as a functional basis of consciousness. *Neuroscience of Consciousness*, 2019(1). <https://doi.org/10.1093/nc/niz016>
- Kiefer, A. B. (2017). Literal perceptual inference. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing*. MIND Group.
- Klein, C., & Barron, A. B. (2020). How experimental neuroscientists can fix the hard problem of consciousness. *Neuroscience of Consciousness*, 2020(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), 307–321. <https://doi.org/10.1038/nrn.2016.22>
- Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences*, 11(1), 16–22. <https://doi.org/10.1016/j.tics.2006.10.012>
- Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & Lange, F. P. de. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex*, 22(9), 2197–2206. <https://doi.org/10.1093/cercor/bhr310>
- Kok, P., Rait, L. I., & Turk-Browne, N. B. (2019). Content-based dissociation of hippocampal involvement in prediction. *Journal of Cognitive Neuroscience*, 32(3), 527–545. https://doi.org/10.1162/jocn_a_01509
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marín, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, 93(3), 480–490. <https://doi.org/10.1016/j.neuron.2016.12.041>
- Lamme, V. A. F. (2020). Visual functions generating conscious seeing. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00083>
- Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, 1(3), 204–220. <https://doi.org/10.1080/17588921003731586>
- Lange, F. P. de, Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, 22(9), 764–779. <https://doi.org/10.1016/j.tics.2018.06.002>
- Lau, H. C. (2008). A higher order Bayesian decision theory of consciousness. *Elsevier*, 168, 35–48. [https://doi.org/10.1016/S0079-6123\(07\)68004-2](https://doi.org/10.1016/S0079-6123(07)68004-2)
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15(8), 365–373. <https://doi.org/10.1016/j.tics.2011.05.009>
- 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>
- Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*, 20(9), 1293–1299. <https://doi.org/10.1038/nn.4615>
- Lettvin, J. Y. (1976). On seeing sidelong. *The Sciences*, 16(4), 10–20. <https://doi.org/10.1002/j.2326-1951.1976.tb01231.x>
- Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64(4), 354–361. <https://doi.org/10.1111/j.1468-0114.1983.tb00207.x>
- Li, H.-H., & Ma, W. J. (2020). Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. *Nature Communications*, 11(1), 2004. <https://doi.org/10.1038/s41467-020-15581-6>
- Limanowski, J., & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00547>
- Litwin, P., & Miłkowski, M. (2020). Unification by fiat: Arrested development of predictive processing. *Cognitive Science*, 44(7), e12867. <https://doi.org/10.1111/cogs.12867>
- Lush, P., Botan, V., Scott, R. B., Seth, A. K., Ward, J., & Dienes, Z. (2020). Trait phenomenological control predicts experience of mirror synaesthesia and the rubber hand illusion. *Nature Communications*, 11(1), 4853. <https://doi.org/10.1038/s41467-020-18591-6>
- Marchi, F., & Hohwy, J. (2020). The intermediate scope of consciousness in the predictive mind. *Erkenntnis*. <https://doi.org/10.1007/s10670-020-00222-7>
- Markov, N. T., Ercsey-Ravasz, M., Essen, D. C. V., Knoblauch, K., Toroczkai, Z., & Kennedy, H. (2013). Cortical high-density counterstream architectures. *Science*, 342(6158). <https://doi.org/10.1126/science.1238406>
- Marshel, J. H., Kim, Y. S., Machado, T. A., Quirin, S., Benson, B., Kadmon, J., Raja, C., Chibukhchyan, A., Ramakrishnan, C., Inoue, M., Shane, J. C., McKnight, D. J., Yoshizawa, S., Kato, H. E., Ganguli, S., & Deisseroth, K. (2019). Cortical layer-specific critical dynamics triggering perception. *Science*, 365(6453). <https://doi.org/10.1126/science.aaw5202>

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, 1(II), 3. <https://doi.org/10.33735/phimisci.2020.II.64>



- Martin, J.-R., & Pacherie, E. (2019). Alterations of agency in hypnosis: A new predictive coding model. *Psychological Review*, 126(1), 133–152. <https://doi.org/10.1037/rev0000134>
- Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, 105(5), 776–798. <https://doi.org/10.1016/j.neuron.2020.01.026>
- Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, 5. <https://doi.org/10.3389/fnhum.2011.00039>
- Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the hierarchical Gaussian filter. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00825>
- McFadden, J. (2020). Integrating information in the brain's EM field: The cemi field theory of consciousness. *Neuroscience of Consciousness*, 2020(1). <https://doi.org/10.1093/nc/niaa016>
- McFadden, J. (2002). The conscious electromagnetic information (Cemi) field theory: The hard problem made easy? *Journal of Consciousness Studies*, 9(8), 45–60.
- Meijs, E. L., Slagter, H. A., Lange, F. P. de, & Gaal, S. van. (2018). Dynamic interactions between top-down expectations and conscious awareness. *Journal of Neuroscience*, 38(9), 2318–2327. <https://doi.org/10.1523/JNEUROSCI.1952-17.2017>
- Melloni, L., Schwiedrzik, C. M., Müller, N., Rodriguez, E., & Singer, W. (2011). Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. *Journal of Neuroscience*, 31(4), 1386–1396. <https://doi.org/10.1523/JNEUROSCI.4570-10.2011>
- Mendonça, D., Curado, M., & Gouveia, S. (2020). *The philosophy and science of predictive processing*. Bloomsbury Publishing.
- Metzinger, T. (2004). *Being no one*. MIT Press.
- Metzinger, T. (Ed.). (2000). *Neural correlates of consciousness: Empirical and conceptual questions*. MIT Press.
- Metzinger, T., & Wiese, W. (Eds.). (2017). *Philosophy and predictive processing*. MIND Group. <https://predictive-mind.net/papers>
- Miller, S. M. (2015). *The constitution of phenomenal consciousness*. John Benjamins.
- 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>
- Millidge, B., Tschantz, A., & Buckley, C. L. (2020). *Whence the expected free energy?* <https://doi.org/10.1101/2004.08128v3>
- Milliere, R., & Metzinger, T. (2020). Radical disruptions of self-consciousness: *Philosophy and the Mind Sciences*, 1(I), 1–13. <https://doi.org/10.33735/phimisci.2020.I.50>
- Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., Goebel, R., & Yacoub, E. (2015). Contextual feedback to superficial layers of V1. *Current Biology*, 25(20), 2690–2695. <https://doi.org/10.1016/j.cub.2015.08.057>
- 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>
- Noë, A. (2004). *Action in perception*. MIT Press.
- Noë, A., & O'Regan, J. K. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939–973.
- Noë, A., & Thompson, E. (2004). Are there neural correlates of consciousness? *Journal of Consciousness Studies*, 11(1), 3–28.
- Noreika, V., Canales-Johnson, A., Harrison, W. J., Johnson, A., Arnatkevičiūtė, A., Koh, J., Chennu, S., & Bekinschtein, T. A. (2017). Wakefulness state modulates conscious access: Suppression of auditory detection in the transition to sleep. *bioRxiv*, 155705. <https://doi.org/10.1101/155705>
- O'Brien, G., & Opie, J. (1999). A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences*, 22, 127–148.
- Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLOS Computational Biology*, 10(5), e1003588. <https://doi.org/10.1371/journal.pcbi.1003588>
- Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, 107(1), 179–217. <https://doi.org/10.1016/j.cognition.2007.09.003>
- Pal, D., Li, D., Dean, J. G., Brito, M. A., Liu, T., Fryzel, A. M., Hudetz, A. G., & Mashour, G. A. (2020). Level of consciousness is dissociable from electroencephalographic measures of cortical connectivity, slow oscillations, and complexity. *Journal of Neuroscience*, 40(3), 605–618. <https://doi.org/10.1523/JNEUROSCI.1910-19.2019>
- Parr, T., Corcoran, A. W., Friston, K. J., & Hohwy, J. (2019). Perceptual awareness and active inference. *Neuroscience of Consciousness*, 2019(1). <https://doi.org/10.1093/nc/niz012>
- Parr, T., & Friston, K. J. (2018). The anatomy of inference: Generative models and brain structure. *Frontiers in Computational Neuroscience*, 12. <https://doi.org/10.3389/fncom.2018.00090>
- Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, 292(5516), 510–512. <https://doi.org/10.1126/science.1057099>
- Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, 1(II), 3. <https://doi.org/10.33735/phimisci.2020.II.64>



- Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge University Press.
- Perrykkad, K., & Hohwy, J. (2020). Fidgeting as self-evidencing: A predictive processing account of non-goal-directed action. *New Ideas in Psychology*, 56, 100750. <https://doi.org/10.1016/j.newideapsych.2019.100750>
- Petzschner, F. H., Weber, L. A., Wellstein, K. V., Paolini, G., Do, C. T., & Stephan, K. E. (2019). Focus of attention modulates the heartbeat evoked potential. *NeuroImage*, 186, 595–606. <https://doi.org/10.1016/j.neuroimage.2018.11.037>
- Pinto, Y., Gaal, S. van, Lange, F. P. de, Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, 15(8), 13–13. <https://doi.org/10.1167/15.8.13>
- Podvalny, E., Yeagle, E., Mégevand, P., Sarid, N., Harel, M., Chechik, G., Mehta, A. D., & Malach, R. (2017). Invariant temporal dynamics underlie perceptual stability in human visual cortex. *Current Biology*, 27(2), 155–165. <https://doi.org/10.1016/j.cub.2016.11.024>
- Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science*, 357(6351), 596–600. <https://doi.org/10.1126/science.aan3458>
- Prinz, J. J. (2012). *The conscious brain: How attention engenders experience*. Oxford University Press.
- Rahnev, D., & Denison, R. N. (2018). Behavior is sensible but not globally optimal: Seeking common ground in the optimality debate. *Behavioral and Brain Sciences*, 41. <https://doi.org/10.1017/S0140525X18002121>
- Reardon, S. (2019). “Outlandish” competition seeks the brain’s source of consciousness. *Science*. <https://www.sciencemag.org/news/2019/10/outlandish-competition-seeks-brain-s-source-consciousness>
- Revonsuo, A. (2006). *Inner presence: Consciousness as a biological phenomenon*. MIT Press.
- Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, 49(3), 329–359.
- Rosenthal, D. M. (1997). A theory of consciousness. In N. Block, O. J. Flanagan, & G. Guzeldere (Eds.), *The nature of consciousness* (pp. 729–753). MIT Press.
- Rudrauf, D., Bennequin, D., Granic, I., Landini, G., Friston, K. J., & Williford, K. (2017). A mathematical model of embodied consciousness. *Journal of Theoretical Biology*, 428, 106–131. <https://doi.org/10.1016/j.jtbi.2017.05.032>
- Sandved Smith, L., Hesp, C., Lutz, A., Mattout, J., Friston, K. J., & Ramstead, M. (2020). Towards a formal neurophenomenology of metacognition: Modelling meta-awareness, mental action, and attentional control with deep active inference. In *PsyArXiv*. <https://doi.org/10.31234/osf.io/5jh3c>
- Schartner, M. M., Carhart-Harris, R. L., Barrett, A. B., Seth, A. K., & Muthukumaraswamy, S. D. (2017). Increased spontaneous MEG signal diversity for psychoactive doses of ketamine, LSD and psilocybin. *Scientific Reports*, 7(1), 46421. <https://doi.org/10.1038/srep46421>
- Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology*, 20(2), 172–176. <https://doi.org/10.1016/j.conb.2010.02.010>
- Searle, J. R. (2000). Consciousness. *Annual Review of Neuroscience*, 23(1), 557–578. <https://doi.org/10.1146/annurev.neuro.23.1.557>
- Seth, A. (2009). Explanatory correlates of consciousness: Theoretical and computational challenges. *Cognitive Computation*, 1(1), 50–63. <https://doi.org/10.1007/s12559-009-9007-x>
- Seth, A. K. (2016). The real problem. *Aeon*. <https://aeon.co/essays/the-hard-problem-of-consciousness-is-a-distraction-from-the-real-one>
- Seth, A. K. (2021). *Being you*. Faber & Faber.
- Seth, A. K. (2015a). Inference to the best prediction. In T. K. Metzinger & J. M. Windt (Eds.), *Open mind*. MIND Group.
- Seth, A. K. (2015b). The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. In T. K. Metzinger & J. M. Windt (Eds.), *Open mind*. MIND Group.
- Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, 5(2), 97–118. <https://doi.org/10.1080/17588928.2013.877880>
- Seth, A. K. (2019). From unconscious inference to the beholder’s share: Predictive perception and human experience. *European Review*, 27(3), 378–410. <https://doi.org/10.1017/S1062798719000061>
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565–573. <https://doi.org/10.1016/j.tics.2013.09.007>
- Seth, A. K., Millidge, B., Buckley, C. L., & Tschantz, A. (2020). Curious inferences: Reply to sun and firestone on the dark room problem. *Trends in Cognitive Sciences*, 24(9), 681–683. <https://doi.org/10.1016/j.tics.2020.05.011>
- Seth, A. K., & Tsakiris, M. (2018). Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences*, 22(11), 969–981. <https://doi.org/10.1016/j.tics.2018.08.008>
- Sherman, M. T., Fountas, Z., Seth, A. K., & Roseboom, W. (2020). Accumulation of salient perceptual events predicts subjective time. *bioRxiv*, 2020.01.09.900423. <https://doi.org/10.1101/2020.01.09.900423>

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, 1(II), 3. <https://doi.org/10.33735/phimisci.2020.II.64>



- Sherman, M. T., Seth, A. K., Barrett, A. B., & Kanai, R. (2015). Prior expectations facilitate metacognition for perceptual decision. *Consciousness and Cognition*, 35, 53–65. <https://doi.org/10.1016/j.concog.2015.04.015>
- Skewes, J. C., Jegindø, E.-M., & Gebauer, L. (2015). Perceptual inference and autistic traits. *Autism*, 19(11). <https://doi.org/10.1177/1362361313519872>
- Skora, L., Seth, A., & Scott, R. B. (2020). Sensorimotor predictions shape reported conscious visual experience in a breaking continuous flash suppression task. In *PsyArXiv*. <https://doi.org/10.31234/osf.io/hwkdj>
- Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A. E., Paliwal, S., Gard, T., Tittgemeyer, M., Fleming, S. M., Haker, H., Seth, A. K., & Petzschner, F. H. (2016). Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience*, 10. <https://doi.org/10.3389/fnhum.2016.00550>
- Stuke, H., Weilhhammer, V. A., Sterzer, P., & Schmack, K. (2019). Delusion proneness is linked to a reduced usage of prior beliefs in perceptual decisions. *Schizophrenia Bulletin*, 45(1), 80–86. <https://doi.org/10.1093/schbul/sbx189>
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9), 403–409. <https://doi.org/10.1016/j.tics.2009.06.003>
- Sun, Z., & Firestone, C. (2020a). The dark room problem. *Trends in Cognitive Sciences*, 24(5), 346–348. <https://doi.org/10.1016/j.tics.2020.02.006>
- Sun, Z., & Firestone, C. (2020b). Optimism and pessimism in the predictive brain. *Trends in Cognitive Sciences*, 24(9), 683–685. <https://doi.org/10.1016/j.tics.2020.06.001>
- Suzuki, K., Schwartzman, D. J., Augusto, R., & Seth, A. K. (2019). Sensorimotor contingency modulates breakthrough of virtual 3D objects during a breaking continuous flash suppression paradigm. *Cognition*, 187, 95–107. <https://doi.org/10.1016/j.cognition.2019.03.003>
- Teufel, C., & Fletcher, P. C. (2020). Forms of prediction in the nervous system. *Nature Reviews Neuroscience*, 21(4), 231–242. <https://doi.org/10.1038/s41583-020-0275-5>
- Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, 21(4), 753–759. [https://doi.org/10.1016/S0896-6273\(00\)80592-9](https://doi.org/10.1016/S0896-6273(00)80592-9)
- Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7), 450–461. <https://doi.org/10.1038/nrn.2016.44>
- Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science*, 282(5395), 1846–1851. <https://doi.org/10.1126/science.282.5395.1846>
- Tononi, G., & Koch, C. (2008). The neural correlates of consciousness. *Annals of the New York Academy of Sciences*, 1124(1), 239–261. <https://doi.org/10.1196/annals.1440.004>
- Tsakiris, M., & Preester, H. D. (2018). *The interoceptive mind: From homeostasis to awareness*. Oxford University Press.
- Tschantz, A., Millidge, B., Seth, A. K., & Buckley, C. L. (2020). *Reinforcement learning through active inference*. <https://arxiv.org/abs/2002.12636>
- Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). Learning action-oriented models through active inference. *PLOS Computational Biology*, 16(4), e1007805. <https://doi.org/10.1371/journal.pcbi.1007805>
- Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. F. (2015). No-report paradigms: Extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences*, 19(12), 757–770. <https://doi.org/10.1016/j.tics.2015.10.002>
- Van de Cruys, S., Friston, K. J., & Clark, A. (2020). Controlled optimism: Reply to Sun and Firestone on the dark room problem. *Trends in Cognitive Sciences*, 24(9), 680–681. <https://doi.org/10.1016/j.tics.2020.05.012>
- Van der Helm, P. A. (2016). Structural coding versus free-energy predictive coding. *Psychonomic Bulletin & Review*, 23(3), 663–677. <https://doi.org/10.3758/s13423-015-0938-9>
- Van Doorn, G., Paton, B., Howell, J., & Hohwy, J. (2015). Attenuated self-tickle sensation even under trajectory perturbation. *Consciousness and Cognition*, 36, 147–153. <https://doi.org/10.1016/j.concog.2015.06.016>
- Van Es, T. (2020). Living models or life modelled? On the use of models in the free energy principle. *Adaptive Behavior*. <https://doi.org/10.1177/1059712320918678>
- Van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, 92(7), 345–381. <https://doi.org/10.2307/2941061>
- Vasser, M., Vuillaume, L., Cleeremans, A., & Aru, J. (2019). Waving goodbye to contrast: Self-generated hand movements attenuate visual sensitivity. *Neuroscience of Consciousness*, 2019(1). <https://doi.org/10.1093/nc/niy013>
- Vogel, D. H. V., Beeker, T., Haidl, T., Kupke, C., Heinze, M., & Vogeley, K. (2019). Disturbed time experience during and after psychosis. *Schizophrenia Research: Cognition*, 17, 100136. <https://doi.org/10.1016/j.scog.2019.100136>
- Vogel, D. H. V., Falter-Wagner, C. M., Schoofs, T., Krämer, K., Kupke, C., & Vogeley, K. (2020). Flow and structure of time experience – concept, empirical validation and implications for psychopathology. *Phenomenology and the Cognitive Sciences*, 19(2), 235–258. <https://doi.org/10.1007/s11097-018-9573-z>

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, 1(II), 3. <https://doi.org/10.33735/phimisci.2020.II.64>



- Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, 1464(1), 242–268. <https://doi.org/10.1111/nyas.14321>
- Weilnhammer, V., Fritsch, M., Chikermane, M., Eckert, A.-L., Kanthak, K., Stuke, H., Kaminski, J., & Sterzer, P. (2020). Evidence for an active role of inferior frontal cortex in conscious experience. *bioRxiv*. <https://doi.org/10.1101/2020.05.28.114645>
- Weilnhammer, V., Stuke, H., Hesselmann, G., Sterzer, P., & Schmack, K. (2017). A predictive coding account of bistable perception – a model-based fMRI study. *PLoS Computational Biology*, 13(5), e1005536. <https://doi.org/10.1371/journal.pcbi.1005536>
- Whyte, C. J. (2019). Integrating the global neuronal workspace into the framework of predictive processing: Towards a working hypothesis. *Consciousness and Cognition*, 73, 102763. <https://doi.org/10.1016/j.concog.2019.102763>
- Whyte, C. J., & Smith, R. (2020). The predictive global neuronal workspace: A formal active inference model of visual consciousness. *bioRxiv*. <https://doi.org/10.1101/2020.02.11.944611>
- Wiese, W. (2018). Toward a mature science of consciousness. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00693>
- Wiese, W. (2015). Perceptual presence in the Kuhnian-Popperian Bayesian brain. In T. K. Metzinger & J. M. Windt (Eds.), *Open mind*. MIND Group.
- Wiese, W. (2020). The science of consciousness does not need another theory, it needs a minimal unifying model. *Neuroscience of Consciousness*, 2020(1). <https://doi.org/10.1093/nc/niaa013>
- Williford, K., Bennequin, D., Friston, K. J., & Rudrauf, D. (2018). The projective consciousness model and phenomenal selfhood. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.02571>
- Woodward, J. (2003). *Making things happen*. Oxford University Press.
- Yon, D., Gilbert, S. J., Lange, F. P. de, & Press, C. (2018). Action sharpens sensory representations of expected outcomes. *Nature Communications*, 9(1), 4288. <https://doi.org/10.1038/s41467-018-06752-7>
- Yon, D., Lange, F. P. de, & Press, C. (2019). The predictive brain as a stubborn scientist. *Trends in Cognitive Sciences*, 23(1), 6–8. <https://doi.org/10.1016/j.tics.2018.10.003>

Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted 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.

