PhiMiSci
Philosophy and the Mind Sciences

You can't always get what you want

Predictive processing and consciousness

Tobias Schlicht[a] (tobias.schlicht@rub.de)
Krzysztof Dołęga[a] (krzysztof.dolega@rub.de)

Abstract
The predictive processing framework has gained significant popularity across disciplines investigating the mind and brain. In this article we critically examine two of the recently made claims about the kind of headway that the framework can make in the neuroscientific and philosophical investigation of consciousness. Firstly, we argue that predictive processing is unlikely to yield significant breakthroughs in the search for the neural correlates of consciousness as it is still too vague to individuate neural mechanisms at a fine enough scale. Despite its unifying ambitions, the framework harbors a diverse family of competing computational models which rely on different assumptions and are under-constrained by neurological data. Secondly, we argue that the framework is also ill suited to provide a unifying theory of consciousness. Here, we focus on the tension between the claim that predictive processing is compatible with all of the leading neuroscientific models of consciousness with the fact that most attempts explaining consciousness within the framework rely heavily on external assumptions.

Keywords
Bayesian brain hypothesis · Consciousness · Instrumentalism · Overflow · Predictive processing · Realism

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

1 Introduction

The predictive processing (PP) framework has caused a lot of excitement among philosophers and cognitive scientists in the last decade. Its main proponents take it to provide a "unified account" of brain function (Friston, 2010), integrating recent developments in computational neuroscience with the insights of Bayesian

[a]Ruhr-University Bochum, Institute of Philosophy II

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80
 https://philosophymindscience.org ISSN: 2699-0369

psychology. Moreover, it is allegedly "the first truly unifying account of perception, cognition and action" (Clark, 2016, p. 2), having the potential to "explain perception and action and everything mental in between" (Hohwy, 2013, p. 1). This is because accounts of cognitive phenomena formulated within this framework posit only one fundamental type of information processing, i.e., prediction error minimization, reiterated throughout the brain on many levels in a complex computational hierarchy. The core idea is that the brain harbours an internal model of the world, the cognitive agent, and the agent's biological interior milieu, which constantly generates expectations (or guesses) about incoming stimuli and their sources on all levels of the hierarchy. These top-down predictions about (the causes of) sensory input are constantly compared to the actual incoming sensory inputs, yielding (bigger or smaller) prediction errors. The internal model and world can be matched in one of two ways: either the input is used to update the model that yielded the predictions in the first place (perceptual inference) or the sensory input is changed via action in order to match the model (active inference).[1] Proponents of such approaches have already demonstrated its strong potential to accommodate various mental phenomena surrounding perception, cognitive penetration, perceptual binding, and attention (cf. Hohwy, 2013, pp. 101–138).

In line with their impressive prior contributions to PP, Hohwy and Seth (2020) as well as Clark, Friston, and Wilkinson (2019), and Solms (2021) among others, have recently stressed the potential of this framework to address consciousness. Hohwy and Seth issued a programmatic statement trying to make good on the claim that predictive processing can also inform us about different facets of conscious experience and guide the search for systematic neural correlates of consciousness (NCCs). They suggest that PP offers the most promising approach for embedding the on-going search for the NCCs within a unifying framework, one which can even motivate and operationalize "closer links between phenomenological properties of conscious experience and mechanistic properties of underlying neural substrates" (Hohwy & Seth, 2020, p. 25). Although they admit that the "PP approach is not, in itself, a theory of consciousness" (Hohwy & Seth, 2020, p. 2), they claim that it is this feature that makes it uniquely fit for providing unification, while remaining optimistic for "the prospect that, eventually, aspects of PP may themselves coalesce into a theory of consciousness in its own right" (Hohwy & Seth, 2020, p. 24). Hohwy and Seth's enthusiasm for PP's ability to facilitate consciousness research echoes that of authors like Clark, Friston, and Wilkinson (2019) as well as Clark (2019), who have proposed ways in which PP can contribute to the philosophical side of explaining consciousness. In particular, they address the so-called hard- and meta-problems of consciousness (Chalmers, 1996, and 2018

[1] Although perceptual and active inference are often distinguished in the literature, the former is formally subsumed by the latter in more recent versions of the framework which assume that the system is minimizing its long-term expected free energy. Minimizing this quantity is taken to be equivalent to maximizing expected utility while simultaneously reducing the uncertainty about the possible causes of valuable outcomes (cf. Friston et al., 2015).

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

respectively) which are concerned with the subjective qualities of experience (commonly referred to as qualia) and the judgements we make and intuitions we have about consciousness. Similarly, Solms (2021) commits explicitly to Friston's free-energy principle in his defense of a "free energy theory of consciousness" (2021, p. 2), a variant of the PP framework, but combines it with suggestions from his work on Freudian neuro-psychoanalysis.

However, it is still unclear whether PP can deliver on all of its promises. While we share much of Hohwy and Seth's optimism, we are also keenly aware of the framework's many limitations as well as outstanding problems, which are often downplayed by its proponents. For example, PP's generality and concomitant compatibility with different existing theories of consciousness could be seen as an obstacle for making a substantial contribution to any specific debate in this area, rather than as an advantage. As we will show, in order to make such contributions, PP must be enriched with further, often externally motivated, assumptions which raise the question whether it is the framework itself or these additional postulates that carry the explanatory burden.

In this paper, we will critically evaluate the explanatory potential of the PP framework with respect to determining the NCCs as well as providing an account of conscious experience more generally. Some of our points are general concerns about the features and the explanatory status of the framework, others are directed at specific variants of PP and can be seen as direct responses to what we see as the most formidable contributions made by Hohwy and Seth, Clark et al., and others. We assume the reader to be familiar with the basics of this explanatory framework for reasons of space.[2] Nevertheless, we begin the paper by introducing the NCC research program and rehearsing the main points of Hohwy and Seth's contribution, in order to clarify the stakes of the debate about NCCs as well as to ensure we do not misrepresent their position (section 2). The remainder of the paper is then divided into two general parts. In section 3, we turn to the general issue of scientific realism with respect to the Bayesian brain, discuss methodological issues surrounding the PP framework, and call into question its supremacy over competing modelling frameworks. Section 4 turns to the philosophical issues surrounding PP and consciousness. First, we point to the fact that the framework cannot answer the hard problem without making additional assumptions. We then evaluate the legitimacy of these assumptions, showing that although the framework is compatible with multiple competing theories of consciousness, the assumptions which allow for this are either motivated by considerations external to the framework or tend to lead to puzzling predictions about phenomenal experience. This is then illustrated using the familiar example of the Müller-Lyer illusion. In sum, we demonstrate that the optimistic attitude towards the prospects of PP contributing to scientific work on consciousness is quite limited, when focusing only on its central theoretical commitments. Thus, in order to contribute to some of the most

[2]See Metzinger and Wiese (2017) for an accessible overview of its main claims and assumptions.

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

interesting and controversial on-going debates surrounding conscious experience, PP must either be amended with additional assumptions or remain silent.

2 PP and the search for the NCC

2.1 Determining neural correlates of consciousness

Despite many methodological challenges and philosophical objections, one of the core projects of current scientific research on consciousness is the search for the 'neural correlates of consciousness' or NCCs (see Chalmers, 2000). The neural correlates in question are usually "defined as the *minimal neuronal mechanisms jointly sufficient for any specific conscious percept*" (Tononi & Koch, 2008, p. 239; see also Fazekas & Overgaard, 2018; Fink, 2016; and Noë & Thompson, 2004 for sceptical looks at this research program). In practice, the aim of this research program is to find brain states or processes which reliably correlate with different aspects of consciousness, which here is used as an umbrella term covering a diverse range of phenomena related to how we experience the world with us in it.

NCC researchers commonly distinguish between consciousness in the sense of a *global state of vigilance* and *specific conscious experiences* of something in various sensory or cognitive modalities. The former allows an organism to receive sensory stimulation and process information at all and comes in degrees, ranging from full wakefulness via various levels of sleep, anaesthesia, vegetative state, all the way down to deep coma (Dehaene et al., 2006; Tononi et al., 2016; Tononi & Koch, 2008). Consciousness in this sense is sometimes also referred to as 'creature consciousness' (Bayne, 2007; Hohwy, 2009). Researchers have already made substantial progress on this aspect of consciousness and its NCCs (see e.g., Fernández-Espejo & Owen, 2013 for a helpful overview). For example, the state of vigilance has been associated with activity in the brainstem, as one's ability to stay alert diminishes after sustaining damage to that area (Damasio, 2011). The other use of the term 'consciousness' refers to particular experiences such as that of pain, color, objects, and so on. These states of consciousness are distinguished by their (phenomenal) contents and can be considered as modifications of our overall unified conscious state at a particular time (Bayne & Chalmers, 2003; Hohwy, 2009; Searle, 1992). For example, looking outside my open window, I may simultaneously *see* trees in the garden, *hear* birds singing, and *smell* the fresh air, accompanied by a mixture of different *feelings* and *thoughts*. The assumption of the NCC project is that for every conscious percept there will be a corresponding neural correlate such that artificially inducing the same activity in the correlated brain structure will induce the percept while disrupting it will eliminate the percept.

The NCC research program emphasises identifying the *minimal* neuronal mechanisms since presumably not all brain areas or neural processes will play a role in generating conscious experience. For example, although it contains many more neurons than the cerebral cortex, the cerebellum does not seem to contribute

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

to the generation of conscious experience (Tononi & Koch, 2015). Moreover, the areas which *do* contribute, need not necessarily do so in the same way. One goal of this research is thus to determine the neuronal *mechanism* sufficient for producing one particular percept instead of another one. This minimal mechanism is sometimes called the 'core' NCC, and it will always be embedded within a 'total' NCC, i.e., a much more encompassing neural activation pattern that also comprises various other mechanisms, e.g., mechanisms in the brainstem necessary for sustaining global consciousness (or vigilance) as described above (see also Marvan & Polák, 2020). In this way, searching for these more specific mechanisms holds promise, ultimately, to arrive at a mechanistic explanation of consciousness.

2.2 PP and the search for systematic NCCs

In their recent article, Hohwy and Seth follow Chalmers (2000) in arguing that the search for the NCC of any particular content of experience should be interlocked *systematically* with research into the correlates of other aspects of experience, such as levels of consciousness as well as phenomenology. Otherwise, any identified NCC would be *arbitrary*, as merely observing patterns of neural activation does not, by itself, allow for uncovering neural mechanisms underlying consciousness. Merely registering that consecutive activity in different neural structures, say x, y, and z, is correlated with some type of experience would not explain what each of these areas contributes to the experience, or why these and not other neural pathways are responsible for that type of experience. The way forward, in their view, is to rely on a larger unified theory of brain function that could yield better predictions about the connections between brain activity and phenomenal states, since being a systematic NCC boils down to "being systematically guided by theoretical considerations of some sort" (Hohwy & Seth, 2020, p. 4).

In this vein, Hohwy and Seth identify three main challenges of NCC research. The first one is that of relating local NCCs for particular contents to global states of consciousness in terms of the overall functioning of the subject's neural system. The fact that any phenomenal state is only a modification of the subject's overall conscious state must inform the search for the NCCs. The second and related issue is that of teasing apart neural activations which are causally contributing to a given experience from those that merely coincide with it. At any given time, there will be many patterns of neural activation differing in resolution and specificity (see e.g., Aru et al., 2012). Finding the *minimal* relevant process requires making inferences about the causal structure of information processing in the brain. Finally, Hohwy and Seth stress that the search for NCCs should integrate action as causally or constitutively relevant for conscious experience. This constraint is in line with recent enactive and embodied approaches to consciousness and cognition (Kirchhoff & Kiverstein, 2018; see, e.g., Noë, 2005) as well as the methodological need for integrating behavioral measures into neuroscience (Krakauer et al., 2017).

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

In addition to the above challenges, Hohwy and Seth also identify two constraints which any promising framework for guiding the search for systematic NCCs should meet. Even though the theoretical considerations which guide this search should not be specifically tied to a particular theory of consciousness, they should, the authors argue, respect the insights of the current neuroscientific consensus about the phenomenon under investigation. By analyzing multiple prominent neuro-centric theories of consciousness, Hohwy and Seth manage to identify two key features which most views hold to be crucial for conscious experience – uncertainty reduction and top-down signalling.[3] Unsurprisingly, these key features also happen to be the core elements of the PP framework.

While the bulk of Hohwy and Seth's paper is devoted to demonstrating how PP can successfully address the three challenges and fit within their constraints in order to provide a theoretical basis for *systematic* NCCs, we think that the core argument of their paper rests on the assumption that PP is not only a correct description of the functions carried out by the brain, but that it is aimed at uncovering the *actual* neural *mechanisms* (in the sense of Illari & Williamson, 2010) underpinning consciousness. As the authors state in their recent paper: "the language of PP enables new and productive mappings to be drawn between phenomenological descriptions and underlying neurocognitive processes" (Hohwy & Seth, 2020, p. 18). It is this assumption that allows them to argue for the explanatory and predictive power of the framework when it comes to the search for NCCs. Simply put, if PP can uncover the real organization and causal interactions between neural components at different spatial and temporal levels, then it can guide meaningful interventions on the mechanism's components, answering to the problem of inferring which patterns of neural activity are causally relevant to a given experience. Furthermore, Hohwy and Seth argue that by offering a complete framework for describing neural mechanisms, PP provides a useful way to systematize and conceptualize conscious experience in a way that will help to solve remaining problems. They illustrate this point by noting that PP already recognizes the features shared by most theoretical proposals to consciousness (reduction of uncertainty and the importance of top-down signalling mentioned above). Again, the PP approach turns out to not only be well positioned to account for these two features, but also offers a unification of the relevant aspects of different neuroscientific accounts of consciousness in a way not afforded by "any other theoretical framework" (Hohwy & Seth, 2020, p. 24).

In this paper we want to question what we see as the two claims made by Hohwy and Seth which also drive much of the enthusiasm for the contributions

[3]Examples of uncertainty reduction provided by Hohwy and Seth include the evidence accumulation threshold for ignition postulated by the Global Neuronal Workspace (Dehaene, 2014), the signal-to-noise ratio criterion postulated by versions of Higher-Order Theories based on signal detection (Lau, 2008), and the exclusion of possibilities used to define conscious states in the Integrated Information Theory (Oizumi et al., 2014). Similarly, the authors point out that some form of top-down modulation is not only included by the views listed above, but applies also to the Recurrent Processing Theory of consciousness (Lamme, 2006).

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

that the framework can make in explaining consciousness. The first of these claims is the presupposition that PP delivers a mechanistic description of brain function, while the second is the claim that the framework offers a complete basis for developing a mature theory of consciousness. As we see it, the kind of work that PP can do is best understood in terms of providing a high-level conceptual/computational framework, but this kind of abstract description does not supply a detailed enough account of how its algorithms can be implemented in neural tissue to deliver the details necessary for guiding the search for systematic NCCs. Moreover, as we show in the later part of this article, such a high-level conceptualization of neural function can be used for formulating an account of consciousness, but it will also need to rely on externally motivated assumptions to achieve this, and even then it may fail to account for some of the aspects of conscious experience.

3 Bayes in the brain?

Like many of the authors working within PP, Hohwy and Seth are committed to a realist attitude, which can be summarized in the expression that "the brain *is* a Bayesian mechanism" (Hohwy, 2013, p. 25). A crucial element of this account is that the brain is equipped with "a single type of mechanism, reiterated throughout the brain" (Hohwy, 2013, p. 2), which can account for all cognitive functions. This is often (and perhaps erroneously, see Colombo & Hartmann, 2017) thought to be the source of the framework's explanatory power. In its strongest form, the conviction about the unifying and explanatory credentials of the framework is expressed in Friston's claim that "if one looks at the brain as implementing this [i.e., PP] scheme, nearly every aspect of its anatomy and physiology starts to make sense" (2009, p. 293).

However, PP's ontological and explanatory status is one of the more contentious issues in philosophy of mind and cognitive science (see e.g., Aitchison & Lengyel, 2017; Bruineberg et al., 2020; Colombo & Wright, 2017; Heilbron & Chait, 2018). Hohwy and Seth acknowledge this when they point out that "the PP framework can be cast at different levels of abstraction which make different claims about the underlying mechanism" (Hohwy & Seth, 2020, p. 15). They go on to distinguish between 'strong' and 'weak' interpretations of the framework. The former are committed to the brain "*in fact* utilizing generative models and approximate Bayesian inference to accomplish prediction error minimization" (2020, p. 15), while the latter are mere re-descriptions of cognitive systems "*as if* they are utilizing such models and inferential processes, though the underlying physical-causal mechanism may be different" (2020, p. 15, emphasis added).

Unfortunately, this distinction between strong and weak interpretations is not as clear as it might seem at first. On the one hand, any model that recapitulates the joint probability distribution of some observable variable x and a latent variable y can be formally described as a generative model (Bishop, 2006; Ng & Jordan, 2001). On the other hand, psychologically plausible approximations to Bayesian in-

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

ference can be achieved without the calculation of explicit probabilities (Sanborn & Chater, 2016). Hohwy and Seth refine the strong approach to PP by pointing out that their commitment to a general, computational-level description provided by the framework constrains the space of viable process theories that can be "implemented mechanistically" through biologically plausible algorithms, which in the case of PP commonly include predictive coding or gradient descent on variational free energy (Hohwy & Seth, 2020, p. 15). However, this way of elaborating the strong interpretation of PP creates the possibility that the framework might offer an accurate description of the computational function (i.e., the abstract mathematical operation) carried out by the cognitive system (in this case approximate Bayesian inference) and still be wrong about the details of how this function is implemented or approximated by the brain (in this case whether it is done via bottom-up propagation of prediction error). In the next two sections we will show that neither a commitment to a particular algorithm nor a commitment to a more coarse-grained view of PP as a high-level conceptual/computational framework are enough to clear some of the challenges for securing systematic NCCs.

3.1 The issue of fine-grained realism about PP

Recall from section 2.2 that in order to establish a systematic NCC, one needs to have means of directing interventions on the brain in such a way as to tease apart which patterns of neural activity are relevant to a given experience. Thus, systematic NCCs can only be found if we have some form of a blueprint about how PP can be implemented in the brain. However, as we will now show, at least one of the ways in which the framework has been claimed to be realized in the brain suffers from severe problems of underdetermination.

Let us turn to predictive coding (PC). Although PC is a popular algorithmic solution among the framework's proponents, there are many ways in which this kind of algorithm could be realized by neural structures (Spratling, 2017). For example (and without getting into too much detail), the most prevalent version of PC originating with Rao and Ballard (1999) and further developed by Friston (2010) and colleagues assumes that neural networks perform inferences through the iterative process of generating predictions and minimizing prediction error – the error in the reconstruction of the input – by updating the coefficients (i.e., the hypotheses about the possible causes of sensory stimulation) so as to minimize the error of future predictions.[4] Importantly, PC schemes growing out of Rao and Ballard's formulation assume prediction errors to be coded as a difference between the expected and actual inputs, and signalled by inhibitory feedback connections. Such connections, as Spratling explains, are usually assumed to be realized by the axon projections of a sub-population of pyramidal cells. However, the problem is that only a relatively minor number of such connections (10–20%) terminate in

[4]Effectively the network is utilizing a common machine learning technique known as gradient descent on residual error (Bishop, 2006, p. 240).

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

inhibitory neurons, while for the vast majority "the primary targets are other pyramidal cells [...], where cortical feedback has an excitatory (modulatory) effect on response; [...] at least in the short term [...]" (Spratling, 2017, p. 6). Furthermore, Rao and Ballard's algorithm involves several other biologically implausible idealizations, such as allowing for the modelled neurons to have negative firing rates (cf. Spratling, 2017).

To deal with these issues, Spratling has devised a competing PC algorithm that aims to alleviate some of the pitfalls of the Rao and Ballard model by utilizing division rather than subtraction for calculating errors (Spratling, 2008a, 2008b; Spratling et al., 2009). This PC algorithm based on divisive input modulation (we will call it PC-DIM for short) has proven successful in not only being able to account for neural phenomena like biased competition, but has also allowed for simulations as many as tens of millions of neurons with hundreds of billions of connections (cf. Spratling, 2017). However, the biggest and most significant difference between these two families of PC algorithms is that Spratling proposes a different grouping of neural populations into processing stages than the one used by Rao and Ballard. This is the most empirically salient of the changes as it means that the two algorithms will map onto cortical structures differently (see Spratling, 2017). The PC-DIM algorithm is consistent with the data about the neural feedback pathways because it groups neural populations in a way that requires all intercortical connections to be excitatory. This implies that all inhibitory connections projecting from error neurons terminate within the processing stage/cortical area in which they are embedded. It should be noted that, while this complicates the typical PC story – since the inter-stage bottom-up channels are no longer communicating an arithmetic difference between the predictions and the input –, Spratling's algorithm still uses them to communicate the way in which one probability distribution diverges from another, effectively implementing an approximation to Bayesian inference. Thus, even though the PC-DIM algorithm does not utilize the forward pathways to propagate prediction errors, but rather the information about the most likely cause, it is still consistent with the more abstract computational scheme postulated by PP.

The example of Spratling's PC-DIM algorithm is helpful in bringing out the first problem with the underdetermination of fine-grained realism about PP and its posits. It shows that process models employed in the framework allow for different kinds of idealizations, meaning that two competing realizations of the same algorithm that postulate different mappings onto cortical structures and, consequently, different ways in which information flows can be consistent with the same neuroanatomical evidence. However, this is not the only problem of underdetermination that the framework is facing.

Rosa Cao (2020) has recently cautioned that the underdetermination of the algorithmic solutions by the empirical data goes beyond different varieties of the same computational scheme and that the same neuroanatomical evidence might be equally consistent with PP as well as other, more traditional message-passing

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

algorithms. She shows that – as far as the flow of information is concerned – PP models of neural signalling can be "relabelled in traditional, non-predictive terms, with no empirical consequences relevant to existing or future data" (Cao, 2020, p. 517).[5] Labelling a neural activity 'predictive' or not is not constrained by the empirical data gathered by scientists unless further assumptions about the coding scheme or implementation relations connecting neural contents to neural signals can be made (Cao, 2020, p. 522). If we assume that PP proponents take notions like 'model,' 'prediction,' 'error' etc. to be representational, i.e., about something, then what is missing is a specification of how to yield a "mapping from informational contents to measurable neural activities" (Cao, 2020, p. 524). As Cao points out, the way to establish this kind of mapping is by exploiting some degree of correlation (and perhaps structural resemblance, see e.g., Shea, 2018) between the elements of the model and the brain. One commonly used framework is that of Shannon's communication theory under which some neural populations are assumed to act as sources of information for other neural systems which act as receivers. Cao argues that what matters is whether a receiving system can decode and react to that information. However, in order for the receiver to decode information, the probability of a signal given the source needs to be specified. In other words, the receiver needs to know what the prior probability of a signal is, in order to react to it appropriately. Thus, Cao's main point is that both the traditional framework and the predictive framework must specify this set of possible states for the system in question which can then be narrowed down by the incoming signal. But then, "given the prior, sending error and sending stimulus have the same effects from the point of view of the system" (Cao, 2020, p. 526). Therefore, it does not add anything of importance whether we call the statistical information 'error signal' or 'bottom-up input.' In this information-theoretic respect, which is crucial for the impact and originality of the framework (as stressed by Hohwy and Seth), PP is demonstrably equivalent to traditional frameworks, not competitive. Another way of putting it is to say that PP adds a different 'gloss' (in the sense proposed in Egan, 2014, 2020) to the same familiar information-theoretic story.

This conclusion seems to be difficult to swallow for PP realists like Clark, Hohwy, or Seth as it means that proponents of PP "have little new to say about the causal structure of the system" in question (Cao, 2020, p. 517). But if this is the case, then how can PP clear the requirement for systematic NCCs? The two ways in which PP is underdetermined by its implementation present a significant

[5] Spratling (2008a, 2008b) and Issa et al. (2018) make similar points, while Summerfield and de Lange (2014) insist that, under some assumptions, PC and evidence accumulation algorithms can be shown to be formally equivalent. Finally, Aitchison and Lengyel end their analysis of different neural implementations of Bayesian and predictive coding schemes by pointing out that "while predictive coding is an attractive algorithmic idea that accounts for a remarkable range of phenomena, the experimental evidence for it seems inconclusive in the sense that it does not rule out Bayesian inference with a direct variable code, potentially in combination with a variety of non-probabilistic processes including attention and adaptation." (Aitchison & Lengyel, 2017, p. 224).

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

difficulty in determining which neural signals or cell populations should be interpreted as carrying predictions within a hierarchical model and which should be interpreted as prediction errors. This is especially relevant to the neuroscience of consciousness, since part of the controversy surrounding the potential NCCs is whether local recurrent signals (e.g., in the primary visual areas) are sufficient for (visual) consciousness or whether recurrent signals originating in prefrontal cortex must also be involved (Dehaene et al., 2006; Lamme, 2006). In light of the multiplicity of available algorithms as well as Cao's misgivings about the framework's ability to offer specific predictions about the causal organization of the brain, it is unclear whether PP could adjudicate between these competing hypotheses.

3.2 The issue of coarse-grained realism about PP

The issues discussed so far have a direct implication for the ways in which PP can guide investigations into the neurological bases of different cognitive phenomena, including the search for NCCs. However, Hohwy and Seth could claim that even though the details of the framework's algorithmic implementations are vague, a coarse-grained version of PP can still make progress in elucidating systematic NCCs. In this section we will argue that the framework is still deficient even when viewed as a more abstracted description of neural mechanisms.

This problem can be better illustrated with an example of a neuroimaging study on the significance of early visual processing in illusory contour perception done by Kok and de Lange (2014). The study used fMRI and population receptive field mapping to investigate the neural correlates of illusory shape perception and perceptual grouping in the primary visual cortex. The researchers presented twenty subjects with a series of similar visual arrays consisting of four 'Pac-man' shapes or inducers, arranged so that each would be equidistant from two others. During test trial presentations of the arrays, three of the four inducer shapes were aligned to form an illusory *Kanizsa triangle*. These inducing configurations were presented for 500ms, separated by a similarly arranged mask image of full circles displayed over the same time interval. Each of the trial sequences lasted 14.4 seconds, and the same positioning of the illusory triangle was used throughout each series. In control configurations, inducers were not aligning into an illusory shape. Additionally, all trials were controlled for attentional spread and modulation by presenting subjects with one of two peripheral detection tasks which did not overlap with the illusory shapes.

The results of comparing the reconstruction of the neural responses to illusory figure configurations with the responses to the non-inducing condition proved more than surprising. V1 regions corresponding to the surface of the illusory figure showed a significant increase in response to illusion inducing configurations, despite the absence of bottom-up input. Concurrently, the response of V1 regions corresponding to the inducer shapes was suppressed in the presence of the illu-

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

sory shape. Both results were also visible in V2, due to the well-known scaling of receptive field sizes in higher levels of the visual cortex (Kastner et al., 2001). Finally, the fixation point falling within or outside of the illusory shape did not significantly modulate these effects. Similarly, subjects' performance in either of the attention control tasks did not show dependence on the presence or absence of an illusory shape.

Kok and de Lange interpret the results of their study by appealing to the top-down modulation of lower cortical areas by higher-level ones. The importance of feedback connections for early visual areas has already been widely confirmed by empirical research on figure-ground segmentation (Zhang & Heydt, 2010), as well as shape and contour detection (Zhou et al., 2000). Most previous studies have stressed the excitatory role of these connections, postulating that they result from higher-level areas passing down information about features relevant for the grouping process. Although this hypothesis explains the observed increase in response during the illusory figure condition, it fails to explain *why* the responses of V1 areas corresponding to non-absent stimulus (here, the inducers) were inhibited in this and in other studies. To accommodate both of the discussed phenomena, Kok and de Lange turn to an interpretation of the results via the lense of PC algorithms (Friston & Kiebel, 2009; e.g., Rao & Ballard, 1999) which assume that the mismatch between the information carried in the feedback signal (e.g., neural activity where there was no activity predicted, or vice versa) will result in a surge of feed forward 'prediction error' signals.

However, without a clear commitment to a particular fine-grained model of the neural populations, this interpretation falls short of disambiguating different possibilities for how the observed activity fits systematically with the working of the wider visual system as well as the phenomenology of the illusory percept. Kok and de Lange's assessment of their results echoes the observation that it is difficult to pinpoint the 'neural signature' of error signals in sensory neurons, which in turn means that it is not clear whether the suppression of activity in cells with receptive fields corresponding to inducer regions is observed because "(1) they are predicted to be active by higher-order areas, (2) this prediction is violated, or (3) both." (Kok & Lange, 2014, p. 1535). Such uncertainty regarding how to interpret these findings is not just a case of limitations in our empirical methods or lack of additional data (though of course, both issues are present here), but also of the unclear empirical predictions made by the framework.

Kok and de Lange's uncertainty about how to interpret their data even after assuming a coarse-grained reading of PP casts a shadow on Hohwy and Seth's optimism about the ways in which PP can facilitate the search for systematic NCCs. According to them, the framework should offer a high-level description which would allow for relating evidence about different neural correlates and how they correspond to and change with conscious experience. However, it is difficult to see how much new information can be gained from relating different empirical findings in this way, if the interpretations of such findings are rendered ambiguous

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

by the PP framework. Hohwy and Seth can reply that it is precisely by locating different empirical findings in a common framework that we can corroborate the available data and resolve such ambiguities, but this strategy can only take us so far. After all, we have already shown that there is more than one way to implement PP, and that competing algorithmic implementations may lead to different interpretations of the data. In short, different PP algorithms may lead to different systematic NCCs.

Notice that all of the above worries are especially problematic when we consider Hohwy and Seth's claim that the framework can solve the challenge of teasing apart neural activity that causally contributes to particular experiences from activity which merely co-occurs with them (i.e., the second challenge mentioned in section 2.2). As they note, the crux to solving this challenge is having the capacity for "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" (Hohwy & Seth, 2020, p. 17). While we have some worries about PP's ability to effectively capture phenomenology (see next sections), we would like to draw attention to the fact that, given what has been said so far, PP simply does not deliver on the promise of guiding causal interventions in a way that the authors envisage. However, our skepticism about PP's inability to meet the challenge of uncovering the mechanisms behind systematic NCCs does not stem only from the framework's underdetermination by the data, but from a further methodological worry, namely that models of cortical activity formulated within PP are usually heavily reliant on the use of causal analysis methods such as dynamic causal modelling (DCM) (Friston et al., 2003). DCM is a statistical method of analysing neuroimaging data which allows for estimating the causal influence between different brain regions by treating them as coupled dynamical systems. The method relies on constructing possible models of cortical interaction and using Bayesian model selection to fit these models to the available measurements. DCM is in many ways an inspiration and a precursor to PP, but the two should not be confused. The former is closer to what philosophers of science have called a model of data – a way of processing and extracting relevant information from measurement (Suppes, 1962), while the latter is more akin to a theoretic model meant to explain why that configuration of data has been observed (Bailer-Jones, 1999; Hartmann, 1995). Most PP accounts of cognitive function based on neuroimaging data not only rely on the use of DCM and similar methods for validation, but also are often built on the results of prior DCM analysis. What this means is that models of neural processing offered by the framework are fitted to the causal analysis of neural connectivity which is largely independent from the PP story.[6] We think that this worry presents a particular problem for those who would like to avoid

[6]The problem of post-hoc fitting of PP models to causal analyses has recently resulted in a controversy regarding which kinds of neurodynamic models support PP descriptions of cognitive functions (Litwin & Miłkowski, 2020).

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

the problem of underdetermination of particular implementations of PP by claiming that the framework still captures some causal relationships between different brain regions on a less-detailed level of description. Simply put, the problem we see here is that the methods of causal analysis which capture causal relations between different brain areas provide the frame which is then glossed over using PP terminology. Therefore, it is those models and not the framework itself that do most of the heavy lifting with regards to explaining the contributions that activity in different neural populations make to particular conscious experiences. Consequently, it is the causal modelling and not PP that seems to help in the search for systematic NCCs. That being said, PP may, in the long run, offer a refined view of why some neural processes causally contribute to particular changes in consciousness over others; but, for now, most of the explanatory burden is carried by auxiliary methods rather than by the framework itself.

Having dealt with the ontological and methodological issues that PP is facing on the road to elucidating consciousness, it is finally time to turn to some of the most famous problems arising for putative explanations of consciousness. Here we think that PP's very general nature prevents it from solving the hard problem without invoking additional assumptions. However, as we will show such assumptions are usually motivated by reasoning and evidence external to the framework. This, in turn, casts doubt on whether it is PP that does the explanatory work in accounts of consciousness formulated within the framework.

4 PP and the hard problem

One of the crucial conceptual issues faced by all proponents of strong PP who want to tackle the mystery of conscious experience is the hard problem of consciousness. Although Hohwy and Seth do not claim that PP can explicitly attack the hard problem, they do think that the framework can offer a piecemeal approach focusing on the search for NCCs that will eventually make the hard problem disappear. As they note, such a "strategy will deliver greater predictive and explanatory power regarding factors that shape and modulate consciousness," while "a focus on systematic NCCs recognizes that 'consciousness' is not a singular explanatory target" (2020, p. 10). This, in turn, is claimed to also provide a better grasp of the overall structure as well as individual differences in phenomenology.

> 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. (Hohwy & Seth, 2020, p. 11)

While the proposed approach can definitely make inroads into the nature of consciousness, it is doubtful that even a complete knowledge of systematic NCCs will yield a satisfying answer to the hard problem. The reason for this is that having complete knowledge about correlations between neural events and phenomenol-

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

ogy would likely not yield, on its own, an explanation of why the states are correlated in these ways. After all, the hard problem of consciousness is not the problem of understanding *what* psychological states are *correlated* with which physical states, but *why* any physical states should *give rise* to any phenomenal experiences.

It is worth noting that Hohwy and Seth's optimism is striking in light of Hohwy's previous statements about the inroads that PP can make in answering to the hard problem. In his book, Hohwy lays the claim that PP "tracks characteristics of conscious perception nicely: binding, penetrability, reality testing, illusions, inextricably rich experience, and first-person perspective," but "is not intended as a proposal that can explain why perceptual states are phenomenally conscious rather than not." (Hohwy, 2013, pp. 201–202). Indeed, (and as the authors note) there are several proposals on the market, which aim to amend PP to solve this problem. Importantly, the existence of such proposals shows that PP cannot solve the hard problem *unless* it is equipped with auxiliary metaphysical assumptions. But what kind of conceptual and methodological assumptions might those be?

4.1 PP and cognitivism about consciousness

Perhaps the most common proposal in the literature (most famously defended by Clark, 2019; but see Dołęga & Dewhurst, 2019; Dołęga & Dewhurst, 2020) is the idea that PP does not respond to the hard problem directly, but rather does so indirectly, by addressing the meta-problem of consciousness first (Chalmers, 2018).[7] Unlike the hard problem, the meta-problem is concerned with our intuitions and judgements about consciousness. This problem concerns answering the question why a cognitive agent may come to make judgments about conscious experience and/or become puzzled by its subjectivity and mysterious nature, rather than try-

[7] Solms (2021, pp. 238–269) diverges from this mainstream by outlining a theory of consciousness in terms of Friston's free energy principle, maintaining that "in order to solve the hard problem of consciousness, science needs to discern the laws governing the mental function of 'feeling' " (Solms, 2021, p. 265). On his view, "everything springs from a system's drive to exist" (Solms, 2021, p. 268) which is explained in terms of the free energy principle, in particular its contribution to explaining self-organization: "Starting with thermodynamics, we arrive – surprisingly easily – at a qualified and agentic subjectivity, one whose most urgent priorities are weighed for a moment, felt, and then transformed with (one hopes) due circumspection into ongoing action." (Solms, 2021, p. 268) Solms objects to Chalmers' panpsychist suggestion that all information has a phenomenal aspect on the grounds that, if that were the case, we would not need (or get) an explanation of consciousness. Thus, he calls for further constraints on the sort of information that has a phenomenal aspect. These come in the form of information processing and possession of a Markov Blanket which together entail the system under investigation is minimising its own entropy (Solms, 2021, p. 263). Although refreshingly direct and relying heavily on the free energy formulation of the PP framework, Seth (2021) does not accept Solms' overall story, especially his take on the hard problem and his identification of the neural origin of consciousness in the brain stem which diverges significantly from the mainstream view that its spring is in the cortex. This disagreement is interesting in light of our criticism since it illustrates the shortcomings of the PP framework – with its diverse formulations – of guiding the search for systematic NCCs.

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

ing to close the stipulated metaphysical gap between consciousness and the physical world. Proponents of tackling consciousness via the meta-problem are, effectively, following Dennett's (1991, 1996, 2015) approach according to which a satisfactory account of consciousness should only need to explain judgements about qualitative feels, irrespective of whether there is or is not anything that the notion of phenomenal consciousness picks out. Thus, philosophers and scientists following this approach only accept consciousness in the sense of conscious access to information (e.g., Dehaene, 2014).

Clark and colleagues' account of consciousness in PP aims to "take the metaphysical sting out of the quale's tail" (Clark et al., 2019, p. 16) by likening qualia to inferred contents, on a par with our projections of other internal and external causes of our sensory inputs. By likening qualia to inferred contents and rejecting any position on which they would be admitted as *data* to be explained (Chalmers, 2013), Clark in effect sides with Dehaene (2014) and Dennett (1996) in rejecting a notion of phenomenal consciousness that asks for explanation *in addition* to cognitive access to information. Thus, the notion of consciousness that gives rise to the hard problem in the first place does not play a role in Clark's account. That it *seems* otherwise to proponents of phenomenal consciousness is not a problem if the PP account can make plausible why it *should* seem that way. And this is where Clark's PP story marries the familiar illusionist approach to consciousness, adding "engineering, neuroscientific, and information-theoretic flesh to the familiar Dennett-style picture" (Clark, 2019, p. 16; see also Dennett, 1991; Frankish, 2016; Kammerer, 2016; Kammerer, 2018 for details on illusionism). Without making it explicit, Clark's take on PP *assumes* a cognitivist account of consciousness, addressing only cognitively accessible contents that can potentially yield perceptual judgements and lead to action. But the crucial explanatory work here is imported from illusionism, rather than being provided by the PP story itself.

Clark takes it as an advantage that his version of PP is *consistent* with Global Workspace Theory (Dehaene, 2014), Information Integration Theory (Tononi, 2008), and Lamme's (2006) Recurrent Processing approach. Hohwy and Seth make analogous comments in praise of PP's flexibility and compatibility with a variety of otherwise competing theories of consciousness. But these competing theories take different stances (and make differing predictions) on important issues, e.g., on whether or not phenomenal consciousness 'overflows' access, on criteria for consciousness (Block, 2011; Dehaene et al., 2006), and on potentially different NCCs for phenomenality and cognitive access (Block, 2005; Tsuchiya et al., 2015). Thus, rather than taking this alleged compatibility as an advantage of the PP account, we submit that it shows the limitations of PP with respect to its ability to make standalone contributions in this area. PP is compatible with the leading neuroscientific theories of consciousness because it fails to make predictions which would rule out any of these theories.

To illustrate our reservations about PP's limited ability to further debates surrounding the assumption of cognitivism brought on board by Clark and colleagues,

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

let us briefly turn to the debate about overflow that is also concerned with the localization of conscious processes in the brain. Is the NCC to be found in sensory areas 'in the back of' the neocortex or does the minimally sufficient NCC involve (pre-)frontal circuits of the brain, which are also associated with cognition and attention (Block, 2019)? As it turns out, this question is closely related to the *conceptual* question about the relation between phenomenology and cognitive accessibility (Block, 2007, 2011) as well as a *methodological* issue of operationalizing consciousness in a way that would prevent conflating neural correlates of consciousness with correlates of cognitive abilities or unconscious processes (see Schlicht, 2018).

Cognitivists like Dehaene (2014, p. 8) have famously defended the idea that "conscious perception must [...] be evaluated by subjective report, preferably on a trial-by-trial basis" (Dehaene et al., 2006, p. 206). Relying on reports presupposes that the contents of experience can be cognitively accessed and that a subject can exhaustively access and report what she phenomenally experiences. This, however, involves a *conceptual* decision on the part of the experimenter which outright dismisses the issue of whether the capacity of consciousness coincides with (or outstrips) the capacity of the cognitive system that is involved in providing access to and enabling report of conscious information. Relying on reports alone prejudges this empirical issue despite evidence to the contrary (Block, 2007, 2011). Thus, since subjective reports not only involve cognitive access but also other cognitive phenomena like attention, working memory, and possibly metacognition, this approach is in danger of confounding a potential NCC with processes underlying these cognitive phenomena.

To overcome this restriction, several researchers have suggested ways to bypass reports by using no-report paradigms in which neural activity is measured without asking subjects to report what they see, for example (Frässle et al., 2014; Tsuchiya et al., 2015). One such strategy exploits the finding that certain automatic eye movements show a high correlation with conscious reports of perceptual dominance during binocular rivalry studies. In their ingenious study, Frässle et al. (2014) made use of the *optokinetic nystagmus* and pupillary reflex. The nystagmus consists of a slow phase in which the eye follows the stimulus, and a fast phase in which the eye quickly reorients in the opposite direction. We are unaware of these movements but they seem to be highly correlated with the percept rather than just the stimulus. Although subjects' patterns of neural activation were superficially similar to the ones measured in the report condition, the intensity in the no-report condition was somewhat reduced and the predominantly frontal activation was largely missing. The contrast between putative NCCs in the two paradigms is striking. This result suggests that frontal activation typical for report studies (used by Dehaene et al., 2006 and others) is due to the involvement of cognitive access to the information in giving the report, not necessarily with phenomenology. On the one hand, proponents of no-report paradigms rightly suggest that their approach is more promising for determining the NCC without confounding it with the neu-

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

ral mechanisms correlated with cognitive functions. On the other hand, this move is in danger of including neural processes within the NCC that are actually underlying *non-conscious* processing instead of conscious processing (Schlicht, 2018). After all, the objective measures used – although tested regarding their correlation with the conscious percept rather than the stimulus – are reflexes outside voluntary control and outside conscious experience. Yet, they are presumably supported by neural mechanisms enabling them. Tsuchiya et al. (2015) acknowledge this possibility but do not elaborate on how to overcome this challenge.[8]

We bring up this example to illustrate that any reconceptualization of the sides of the debate in terms of PP does not seem to get us any closer to a solution. On the one hand, the generality of PP and its compatibility with Dehaene's Global Workspace, Tononi's Integration Information, and Lamme's Recurrent Processing approach, prevent it from contributing any detail which could adjudicate between these theories. On the other hand, Clark et al.'s (2019) commitment to a cognitivist stance on consciousness is too specific and does not allow proponents of PP to take the possibility of phenomenal overflow seriously.

To drive home this point we now want to look at a different assumption brought in by proponents of PP accounts of consciousness. The final point we wish to make concerns phenomenology, or more specifically, the contents of phenomenal experience. Using a familiar example, we illustrate that PP not only fails to settle empirical debates in research on consciousness, but can also make false predictions about phenomenal experience.

4.2 PP, mid-levels, and phenomenology

According to Hohwy, the PP story implies that "perceptual content *is* the predictions of the currently best hypothesis about the world" (Hohwy, 2013, p. 48). In other words, perceptual content is equated with the content of the probabilistic representations that make up the model of the world constructed by the brain. Clark agrees by claiming that "what we experience as qualia are simply content items in our best generative model of the world [...]; conscious contents emerge as precision-driven best-estimates of organism-salient (potentially action-driving) states of affairs" (Clark, 2019, p. 17, n. 43).

As already mentioned, Clark, Friston, and Wilkinson (2019) aim to dispel concerns about the hard problem by responding to the meta-problem. This response holds that the brain, viewed as a kind of "inference machine will be led to conclude that it is home to some very puzzling states that have many of the hallmarks of 'qualia' " (Clark et al., 2019, p. 20). More specifically, given that the brain is supposed to harbour a multi-layered hierarchical generative model of (the causes of) its sensory inputs, any creature equipped with such a brain, "will represent some of its mid-level inferences as especially certain. These mid-level states confidently re-code raw sensory stimulation in ways that [...] fall short of fully de-

[8] See Block (2019) and Michel and Morales (2020) for further discussion of these issues.

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

termining how properties and states of affairs are arranged in the distal world." (Clark et al., 2019, p. 19). Thus, the authors appeal to the level of *certainty* associated particularly with mid-level hypotheses within the complex hierarchy of top-down and bottom-up processing in the brain. That is, particular qualia are assumed to be identical to inferred causes "that are also represented as especially certain" (Clark et al., 2019, p. 21).[9]

It should be noted that, despite the disagreements between Hohwy's and Clark's interpretations of the PP framework, Hohwy also endorses a similar view (Marchi & Hohwy, 2020), though his way of arriving at this position is somewhat different. From his perspective, what specifies which levels of the hierarchy will be privileged in relation to consciousness is dependent not only on the complexity of models that the organism's cognitive system is capable of harbouring, but also on the range of actions that the organism can engage in its environment. Therefore, what determines the privileged role played by the intermediate levels with regard to human consciousness are the contingent ecological and evolutionary facts about the temporal scales at which humans can act. Thus, on this view, "intermediateness is not an essential feature of consciousness" (Marchi & Hohwy, 2020, p. 1). While Hohwy and Marchi develop this position in greater nuance, what is relevant here is that, like Clark, they defend the intermediate level view of consciousness as particularly important. There are (at least) three problems with this approach.

Firstly, one should ask why the representational states that are important to consciousness should be located on the *mid-level*? Clark et al. (2019) emphasize the amount of certainty, which we allegedly assign to particular mid-level hypotheses (encodings). Why? What is special about them in contrast to hypotheses on other levels? And what does certainty have to do with phenomenology? Why should there be a principled connection, let alone an explanatory one? Generally speaking, all hypotheses in the hierarchical generative model are constrained and determined in two directions, namely, by lower-level sensory feedback and by higher-level hypotheses. This ubiquitous dependency and "entanglement" (Clark, 2019) sheds doubt on singling out a particular level which could be held responsible for the "puzzlement … concerning the 'explanatory gap,' where we are almost fooled into believing that there's something special about qualia – that they are not simply highly certain mid-level encodings" (Clark et al., 2019, p. 24). To evaluate this, it is important to note that 'mid-level' is shorthand for a multiplicity of

[9]Of course, Clark et al.'s (2019) account is more refined. Following Dennett (2013, 2015), Clark elaborates that on the basis of its interaction with the world, the embodied agent's brain is prone to infer qualitative features like cuteness, redness, painfulness etc. as objective (internal or external) features just like it infers the existence of babies, tomatoes, and toothaches as potential causes of sensory input. But the agent does not get cognitive or conscious access to the on-going process of inference, only to the "final estimations" (2019, p. 11), which Clark considers to be adaptively reasonable and economical. To the extent that interoceptive features of the agent's internal bodily milieu (see Damasio, 1999; Seth, 2013) drive unconscious inference to qualia, the inferred worldly objects themselves involve predictions that track our own reactive dispositions.

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

levels in the hierarchy. This raises the question where in the hierarchy exactly the so-called mid-level starts and ends. In which direction is certainty supposed to increase or decrease, in the direction of greater specificity or greater generality?

As it turns out, most of the authors appealing to the mid-level view look outside of the framework to answer these questions. Firstly, they seem to motivate their view either by an appeal to evidence in favour of previously available mid-level accounts (Jackendoff, 1987; Prinz, 2012)[10] or other factors external to PP, like the perspective from the embodied and ecological research programs adopted by Marchi and Hohwy (2020). Thus, the mid-level hypothesis requires not only further motivation from *within* the PP frameworks, but also a more detailed treatment. This is especially important since, as has been previously discussed by Dołęga and Dewhurst (2020), the current globally best hypothesis must be cashed out in terms of many hypotheses pertaining to the multiple levels of the hierarchy, where each of them simultaneously makes predictions about the level below and is constrained by that level. But if higher-level hypotheses constrain and determine lower-level (thus including mid-level) hypotheses, then this leads to a further problem which we illustrate below.

Consider the well-known Müller-Lyer illusion in which two horizontal lines of equal length appear to us as being different in length. Now, we can take our hands (engage in active inference) and cover the arrows that induce the illusion, just to be sure and reduce uncertainty. Once we do that, we will see that the lines are of equal length. However, the illusion returns once we remove our hands.[11] The peculiarity of the Müller-Lyer effect is that it persists even once we *know* that the lines are of equal length.[12] Let's assume then, following Clark et al. (2019) and Hohwy (2013), that 'what it is like' for me to see the lines (appearing different in length) is determined by the currently best guess coded in the hierarchical model. In Clark's case this will be the mid-level hypothesis, which is taken to be especially certain. The problem then is that, intuitively, we should expect that the sensory evidence gathered by covering the arrows or by measuring them (i.e., via active inference), and the subsequent updating, should *increase* the brain's confidence in the belief that they are *equal* in length. After all, this evidence should be taken more seriously than the evidence based on looking alone (i.e., perceptual

[10] These views usually stress the importance of working memory as one of the crucial vehicles for conscious contents. We see this as problematic for PP because no substantial treatment of the nature and role of working memory has been offered within the framework.

[11] Some proponents of PP may object that there is a significant difference between the scenario in which the lines are perceived with the inducers visible and the one where they are covered, since the latter involves an action which effectively changes the visual stimulus and the state of the environment that the system is trying to infer. This is right, but notice that in both cases the perceptual system is inferring the length of the same two lines, i.e., the content of the hypothesis about the length of the lines should be the same after the inducer arrows are made visible.

[12] This has led philosophers to argue for the encapsulation of perception, see Fodor (1983), starting a debate with Churchland (1988) about the penetrability of perception by cognitive processes, which lasts until today (see McCauley & Henrich, 2006; and Nes & Chan, 2020 for helpful overviews and a summary of recent developments).

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

inference). Nevertheless, we continue to experience the lines as being *different* in length, our perceptual faculties seem to be cognitively impenetrable to our perceptual beliefs. While the PP approach that aligns phenomenality with assigned certainty to a given hypothesis predicts that one should experience the lines being equal in length, this does not seem to fit the fact that one remains subject to the illusion that they are unequal in length. If the reasoning above is correct, then the hypothesis that determines phenomenal character (*different in length*) differs from the hypothesis with the highest certainty (*equal in length*). This also creates a problem of defining what it means for a hypothesis (i.e., an internal representation of the external world) to be 'mid-level' in the sense relevant for the meta-problem response relying on the certainty of mid-level hypotheses, further undermining its persuasive strength towards dispelling the hard problem.

5 Conclusion

In this paper, we critically assessed the proposal that the popular PP approach to perception, cognition, and action can be also successfully applied to consciousness and thus guide the search for the neural correlates of consciousness. As we have shown, it is questionable whether or not the PP framework can, in fact, offer a componential analysis of the nervous system which could guide empirical investigations aimed at uncovering the mechanisms underlying conscious experiences. We have also demonstrated that the accounts of consciousness currently utilizing PP are usually dependent on assumptions external to the framework for conceptualizing and accounting for consciousness. Such accounts, furthermore, often lead to additional problems and misleading predictions.

Despite the critical voice of this paper, we do not want to suggest that the PP framework cannot be useful to neuroscientific research into consciousness *on principle*. Our criticism is only meant as a voice of caution which highlights that the hype surrounding the framework often outstrips the support that current research lends to it. It is simply too early in the development of the framework to reliably judge whether or not it can deliver on many of its promises. We should remain cautious about the grand claims about PP's explanatory potential and focus on filling in the details as well as overcoming conceptual and methodological obstacles that the framework is facing.

Acknowledgments
Work on this paper was generously supported by the Volkswagen Foundation (grant no. 87 105 - 1 for our research project 'Situated cognition. Perceiving the world and understanding other minds.') We are grateful for this support and for the helpful comments on an earlier draft by our team members Marco Facchin, Paola Gega, François Kammerer, Nina Poth, Bartosz Radomski, Tobias Starzak and Elmarie Venter, two anonymous reviewers and the editors of this journal.

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

References

Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, *46*, 219–227. https://doi.org/10.1016/j.conb.2017.08.010

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

Bailer-Jones, D. M. (1999). Tracing the development of models in the philosophy of science. In L. Magnani, N. J. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 23–40). Springer US.

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

Bayne, T., & Chalmers, D. J. (2003). What is the unity of consciousness? In A. Cleermans (Ed.), *The unity of consciousness: Binding, integration, dissociation* (pp. 23–58). Oxford University Press.

Bishop, C. (2006). *Pattern recognition and machine learning*. Springer-Verlag.

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

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30*(5-6), 481–499. https://doi.org/10.1017/S0140525X07002786

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, *15*(12), 567–575. https://doi.org/10.1016/j.tics.2011.11.001

Block, N. (2019). What is wrong with the no-report paradigm and how to fix it. *Trends in Cognitive Sciences*, *23*(12), 1003–1013. https://doi.org/10.1016/j.tics.2019.10.001

Bruineberg, J., Dołęga, K., Dewhurst, J., & Baltieri, M. (2020). *The emperor's new Markov blankets*. http://philsci-archive.pitt.edu/18467/

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

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford Univeristy Press.

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

Chalmers, D. J. (2013). How can we construct a science of consciousness? *Annals of the New York Academy of Sciences*, *1303*(1), 25–35. https://doi.org/https://doi.org/10.1111/nyas.12166

Chalmers, D. J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, *25*(9-10), 6–61.

Churchland, P. M. (1988). Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science*, *55*(June), 167–187. https://doi.org/10.1086/289425

Clark, A. (2016). *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., & Wilkinson, S. (2019). Bayesing qualia: Consciousness as inference, not raw datum. *Journal of Consciousness Studies*, *26*(9-10), 19–33.

Colombo, M., & Hartmann, S. (2017). Bayesian cognitive science, unification, and explanation. *The British Journal for the Philosophy of Science*, *68*, 451–484.

Colombo, M., & Wright, C. (2017). Explanatory pluralism: An unrewarding prediction error for free energy theorists. *Brain and Cognition*, *112*, 3–12. https://doi.org/10.1016/j.bandc.2016.02.003

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Harcourt Brace; Co.

Damasio, A. R. (2011). *Self comes to mind: Constructing the conscious brain*. Pantheon.

Dehaene, S. (2014). *Consciousness and the brain*. Viking.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211. https://doi.org/10.1016/j.tics.2006.03.007

Dennett, D. C. (2013). Expecting ourselves to expect: The Bayesian brain as a projector. *The Behavioral and Brain Sciences*, *36*(3), 209–210. https://doi.org/10.1017/S0140525X12002208

Dennett, D. C. (1991). Real patterns. *Journal of Philosophy*, *88*(1), 27–51. https://doi.org/10.2307/2027085

Dennett, D. C. (1996). Facing backwards on the problem of consciousness. *Journal of Consciousness Studies*, *1*(3), 4–6.

Dennett, D. C. (2015). *Why and how does consciousness seem the way it seems?* (W. Wanja & T. Metzinger, Eds.). Open MIND.

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

Dołęga, K., & Dewhurst, J. (2019). Bayesian frugality and the representation of attention. *Journal of Consciousness Studies*, *26*(3-4), 38–63.

Dołęga, K., & Dewhurst, J. (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

Egan, F. (2014). How to think about mental content. *Philosophical Studies*, *170*(1), 115–135. https://doi.org/10.1007/s11098-013-0172-0

Egan, F. (2020). A deflationary account of mental representation. In J. Smortchkova, K. Dołęga, & T. Schlicht (Eds.), *What are mental representations?* (pp. 26–53). Oxford University Press.

Fazekas, P., & Overgaard, M. (2018). Perceptual consciousness and cognitive access: An introduction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1755). https://doi.org/10.1098/rstb.2017.0340

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." In *Frontiers in Psychology* (Vol. 7, p. 1044). https://doi.org/10.3389/fpsyg.2016.01044

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, *23*(11-12), 11–39.

Frässle, S., Sommer, J., Jansen, A., Naber, M., & Einhäuser, W. (2014). Binocular rivalry: Frontal activity relates to introspection and action but not to perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(5), 1738–1747. https://doi.org/10.1523/JNEUROSCI.4403-13.2014

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302. https://doi.org/10.1016/s1053-8119(03)00202-7

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Friston, K., 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

Hartmann, S. (1995). Models as a tool for theory construction: Some strategies of preliminary physics. In W. Herfel, W. Krajewski, I. Niiniluoto, & R. Wójcicki (Eds.), *Theories and models in scientific processes* (pp. 26–53). Rodopi.

Heilbron, M., & Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

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. (2013). *The predictive mind*. Oxford: 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). https://doi.org/10.33735/phimisci.2020.ii.64

Illari, P. M., & Williamson, J. (2010). Function and organization: Comparing the mechanisms of protein synthesis and natural selection. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *41*(3), 279–291. https://doi.org/10.1016/j.shpsc.2010.07.001

Issa, E. B., Cadieu, C. F., & DiCarlo, J. J. (2018). Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. *eLife*, *7*, e42870. https://doi.org/10.7554/eLife.42870

Jackendoff, R. (1987). *Consciousness and the computational mind*. MIT Press.

Kammerer, F. (2016). The hardest aspect of the illusion problem — And how to solve it. *Journal of Consciousness Studies*, *23*(11–12), 124–139.

Kammerer, F. (2018). Can you believe it? Illusionism and the illusion meta-problem. *Philosophical Psychology*, *31*(1), 44–67. https://doi.org/10.1080/09515089.2017.1388361

Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology*, *86*(3), 1398–1411. https://doi.org/10.1152/jn.2001.86.3.1398

Kirchhoff, M. D., & Kiverstein, J. (2018). *Extended consciousness and predictive processing: A third-wave view*. Routledge.

Kok, P., & Lange, F. P. de. (2014). Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Current Biology: CB*, *24*(13), 1531–1535. https://doi.org/10.1016/j.cub.2014.05.042

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, 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. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501. https://doi.org/10.1016/j.tics.2006.09.001

Lau, H. (2008). A higher order Bayesian decision theory of consciousness. In R. Banerjee & B. K. Chakrabarti (Eds.), *Models of Brain and Mind: Physical, Computational, and Psychological Approaches* (pp. 35–48). Elsevier.

Litwin, P., & Miłkowski, M. (2020). Unification by fiat: Arrested development of predictive processing. *Cognitive Science*, *44*(7), e12867. https://doi.org/https://doi.org/10.1111/cogs.12867

Marchi, F., & Hohwy, J. (2020). The intermediate scope of consciousness in the predictive mind. *Erkenntnis*. https://doi.org/10.1007/s10670-020-00222-7

Marvan, T., & Polák, M. (2020). Generality and content-specificity in the study of the neural correlates of perceptual consciousness. *Philosophy and the Mind Sciences*, *1*(II). https://doi.org/10.33735/phimisci.2020.II.61

McCauley, R. N., & Henrich, J. (2006). Susceptibility to the Müller-Lyer illusion, theory-neutral observation, and the diachronic penetrability of the visual input system. *Philosophical Psychology*, *19*(1), 79–101. https://doi.org/10.1080/09515080500462347

Metzinger, T., & Wiese, W. (2017). *Philosophy and predictive processing*. MIND Group.

Michel, M., & Morales, J. (2020). Minority reports: Consciousness and the prefrontal cortex. *Mind & Language*, *35*(4), 493–513. https://doi.org/https://doi.org/10.1111/mila.12264

Nes, A., & Chan, T. (2020). *Inference and consciousness*. London: Routledge.

Ng, A. Y., & Jordan, M. I. (2001). *On discriminative vs. Generative classifiers: A comparison of logistic regression and naive bayes*. 841–848. https://doi.org/10.5555/2980539.2980648

Noë, A. (2005). *Action in perception*. MIT Press.

Noë, A., & Thompson, E. (2004). Are there neural correlates of consciousness? *Journal of Consciousness Studies*, *1*(11), 3–28.

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

Prinz, J. J. (2012). *The conscious brain: How attention engenders experience*. Oxford University Press.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *20*(12), 883–893. https://doi.org/10.1016/j.tics.2016.10.003

Schlicht, T. (2018). A methodological dilemma for investigating consciousness empirically. *Consciousness and Cognition*, *66*, 91–100. https://doi.org/10.1016/j.concog.2018.11.002

Searle, J. (1992). *The rediscovery of the mind*. MIT Press.

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. (2021). *Mixed feelings about a hard problem: Review of the hidden spring*. https://neurobanter.com/2021/02/18/mixed-feelings-about-a-hard-%0Aproblem-review-of-the-hidden-spring

Shea, N. (2018). *Representation in cognitive science*. Oxford University Press.

Solms, M. (2021). *The hidden spring: A journey to the source of consciousness*. London, UK: Profile Books.

Spratling, M. W. (2008a). Predictive coding as a model of biased competition in visual attention. *Vision Research*, *48*(12), 1391–1408. https://doi.org/10.1016/j.visres.2008.03.009

Spratling, M. W. (2008b). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience*, *2*. https://doi.org/10.3389/neuro.10.004.2008

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, *112*, 92–97. https://doi.org/10.1016/j.bandc.2015.11.003

Spratling, M. W., De Meyer, K., & Kompass, R. (2009). Unsupervised learning of overlapping image components using divisive input modulation. *Computational Intelligence and Neuroscience*, *2009*, e381457. https://doi.org/10.1155/2009/381457

Summerfield, C., & Lange, F. P. de. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews. Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Suppes, P. (1962). Models of data. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology and philosophy of science: Proceedings of the 1960 international congress* (pp. 252–261). Stanford University Press.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *The Biological Bulletin*, *215*(3), 216–242. https://doi.org/10.2307/25470707

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80

Tononi, G., Boly, M., Gosseries, O., & Laureys, S. (2016). The neurology of consciousness. In G. Tononi, M. Boly, O. Gosseries, & S. Laureys (Eds.), *The neurology of conciousness* (pp. 407–461).

Tononi, G., & Koch, C. (2008). The neural correlates of consciousness: An update. *Annals of the New York Academy of Sciences*, *1124*, 239–261. https://doi.org/10.1196/annals.1440.004

Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1668). https://doi.org/10.1098/rstb.2014.0167

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

Zhang, N. R., & Heydt, R. von der. (2010). Analysis of the context integration mechanisms underlying figure-ground organization in the visual cortex. *Journal of Neuroscience*, *30*(19), 6482–6496. https://doi.org/10.1523/jneurosci.5168-09.2010

Zhou, H., Friedman, H. S., & Heydt, R. von der. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *20*(17), 6594–6611.

Schlicht, T. & Dołęga, K. (2021). You can't always get what you want: Predictive processing and consciousness. *Philosophy and the Mind Sciences*, *2*, 8. https://doi.org/10.33735/phimisci.2021.80