Fame in the predictive brain: a deflationary approach to explaining consciousness in the prediction error minimization framework

Krzysztof Dołęga¹ · Joe E. Dewhurst²

Received: 12 June 2018 / Accepted: 20 January 2020 / Published online: 29 January 2020
© The Author(s) 2020

Abstract

The proposal that probabilistic inference and unconscious hypothesis testing are central to information processing in the brain has been steadily gaining ground in cognitive neuroscience and associated fields. One popular version of this proposal is the new theoretical framework of predictive processing or prediction error minimization (PEM), which couples unconscious hypothesis testing with the idea of ‘active inference’ and claims to offer a unified account of perception and action. Here we will consider one outstanding issue that still looms large at the core of the PEM framework: the lack of a clear criterion for distinguishing conscious states from unconscious ones. In order to fulfill the promise of becoming a unifying framework for describing and modeling cognition, PEM needs to be able to differentiate between conscious and unconscious mental states or processes. We will argue that one currently popular view, that the contents of conscious experience are determined by the ‘winning hypothesis’ (i.e. the one with the highest posterior probability, which determines the behavior of the system), falls short of fully accounting for conscious experience. It ignores the possibility that some states of a system can control that system’s behavior even though they are apparently not conscious (as evidenced by e.g. blindsight or subliminal priming). What follows from this is that the ‘winning hypothesis’ view does not provide a complete account of the difference between conscious and unconscious states in the probabilistic brain. We show how this problem (and some other related problems) for the received view can be resolved by augmenting PEM with Daniel Dennett’s multiple drafts model of consciousness. This move is warranted by the similar roles that attention and internal competition play in both the PEM framework and the multiple drafts model.

✉ Krzysztof Dołęga
krzysztof.dolega@rub.de

¹ Institut für Philosophie 2, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany

² MCMP, Ludwig-Maximilians-Universität München, Ludwigstraße 31, München 80539, Germany

Keywords Predictive processing · Active inference · Consciousness · Multiple drafts · Prediction error minimization · Generative model · Control model · Attention · Illusionism

1 Introduction

The prediction error minimization framework (henceforth, PEM) aims to offer a unified explanation of the processes underlying perception and cognition by postulating that all cognitive functions can be reduced to one kind of process: minimizing prediction error. It has recently been popularised in the philosophical literature by Hohwy (2013) and Clark (2013, 2016), and has been increasingly gaining traction in both empirically informed philosophy of mind and cognitive (neuro)science. One of the reasons for PEM’s widespread adoption is the claim that the framework can deliver a unified theory of all cognitive functions. In the words of Clark, it is “the first truly unified account of perception, cognition, and action” (Clark 2016, p. 2). Similarly, Hohwy claims that PEM is “meant to explain perception, and action, and everything mental in between” (Hohwy 2013, p. 1). Unsurprisingly, the lofty unificatory and explanatory ambitions of the framework have come under much scrutiny (Colombo and Hartmann 2017; Gładziejewski 2019). However, there has been relatively little focus on the fact that, in order to offer a truly exhaustive account of our mental lives, PEM must also face the problem of explaining the processes and mechanisms underlying consciousness.

Our aim in this paper is to assess the (currently) most popular proposal for accounting for consciousness under the PEM framework and identify problems with its underlying assumptions, which we think can be resolved by applying a deflationary interpretation drawn from the work of Daniel Dennett and other like minded philosophers and scientists (e.g. Dennett 1991; Humphrey 2011; Graziano 2013; Pereboom 2011; Frankish 2016a, b; Dołęga and Dewhurst 2019). Our approach will show how consciousness can be accommodated by the PEM framework, while also resolving some long-standing philosophical issues surrounding that topic.

In Sect. 2 we will introduce some basic aspects of the PEM framework, without going into more detail than is strictly necessary for our argument. These aspects are prediction error (2.1), generative models (2.2), active inference (2.3), and precision optimisation (2.4). We will then present Jakob Hohwy’s proposal that conscious experience is determined by the current winning hypothesis (2.5), before moving on in Sect. 3 to identify some issues with this proposal. We will make explicit a representational assumption lying behind the implicit relationship between the properties of conscious experience and the properties of generative models, which leads to two problems—the problem of unconscious representation (3.1) and the problem of unconscious perception (3.2). As we argue, both of these related issues threaten to trivialize or severely limit the scope of Hohwy’s account of consciousness. In Sect. 4 we will present Daniel Dennett’s approach to studying consciousness and the resulting multiple drafts model (especially the notion of a ‘probe’ that determines the content of consciousness, discussed in 4.3), before arguing that it is a natural fit for the PEM framework, and that it offers a deflationary solution to both problems. One consequence of applying the multiple drafts model to PEM is that it urges the framework’s

proponents to adopt a somewhat revisionary attitude towards our understanding of consciousness, as it will turn out that we may be systematically misled about the nature and richness of our conscious experiences (4.4). We conclude by considering some of the further implications of this approach for the future development of the PEM framework.

2 Basic tenets of the prediction error minimization framework

PEM aims to completely overhaul the way we think about the organization and the functioning of the brain. According to this framework, the brain does not process information by accumulating raw sensory data and processing them into ever more detailed and complex representations of the world (as proposed, for example, by Marr 1982). Rather, it postulates that the nervous system operates by generating anticipatory hypotheses about the environmental causes of its future inputs, which are then tested against the actual states of the sensory surfaces and updated accordingly. This is accomplished by a hierarchically organised structure that is driven in a top-down manner, with each level predicting the activity of the level below.¹ Predictions are made based on a ‘generative model’ of the external world, and cascade down the hierarchy, constraining activity on each of the levels below. These predictions terminate in the sensory periphery, where they are ‘tested’ against sensory inputs, with any error being passed back up the hierarchy in order to correct the model. In this section we will briefly review some of the most important features of the framework,² before considering the leading proposal for how it might try to account for consciousness.

2.1 Prediction error

Unlike in traditional models of cognition and early sensory processing, only information that is inconsistent with the downward flow of predictions (i.e. prediction error) is propagated up the hierarchy for further processing (Rao and Ballard 1999; Friston 2010). This error signal carries information about the divergence between the predicted pattern of sensory stimulation and the actual state of the sensory periphery, and is propagated back up the hierarchy until it reaches a level at which it can be integrated into the system’s future predictions (Clark 2013, p. 187). This process of revising and updating the current generative model can be seen as a form of implicit, on-line learning, and begins as soon as prediction error is fed up the hierarchy. The process takes place over multiple levels and time scales, offering a high degree of plasticity (continuous revision) and efficiency (through its anticipatory nature) in real time processing.

¹ The notion of a ‘level’ is somewhat underdetermined in the PEM literature. It is usually assumed that different portions of the PEM hierarchy will map onto different cortical areas, but the individuation of what counts as a distinct cortical area may differ between particular implementations (Spratling 2017).

² For a more detailed philosophical overview see Wiese and Metzinger (2017), and for a technical tutorial see Bogacz (2017). Here we focus just on presenting the details that are most pertinent to our discussion of consciousness.

2.2 Generative models

One of the notions which lie at the heart of the PEM proposal is that of a generative model. Clark (2013) summarizes the role that these models play in the framework as follows:

A generative model [...] aims to capture the statistical structure of some set of observed inputs by tracking (one might say, by schematically recapitulating) the causal matrix responsible for that very structure. [...] In practice, this means that top-down predictions within a multi-level (hierarchical and bidirectional) system come to encode a probabilistic model of the activities of units and groups of units within lower levels, thus tracking [...] interacting causes in the signal source, which might be the body, or the external world [...] (Clark 2013, p. 182).

Therefore, the system not only learns the regularities in the sensory input, but also infers abstract rules governing such regularities. By building a statistical model of the environment and capturing the structure of the underlying ‘causal matrix’ (Clark 2013, p. 182) responsible for the sensory data the system can efficiently predict its input. This can be done by a variety of statistical methods such as, for example, ‘empirical Bayes’, in which priors (i.e. the statistical ‘beliefs’ or expectations which form the basis for the top-down predictions) are initially extracted from the regularities in the bottom-up signal and refined with consecutive inferences (Friston and Kiebel 2009), or variational Bayesian methods (e.g. Mathys et al. 2011, 2014), in which the problem of inferring the true causes of sensory stimulation is reformulated as one of optimizing the parameters of the internal model so as to minimize some objective function (in this case the divergence between the predicted and actual state of the sensory periphery). The exact method used does not matter for present purposes, although ultimately a plausible mechanism will have to be proposed for its neural implementation (cf. Milkowski 2016).

2.3 Active inference

Passively performing inferences and updating the generative model is, for many purposes, not enough to accommodate incoming prediction error. As Hohwy notes: “a system without agency cannot minimize surprise but only optimize its models of the world” (2012, p. 3). Thus, the system needs to also perform *active* inference (Clark 2013; Friston et al. 2009; Hohwy 2012), in which it not only predicts the changes in its own states, but also engages in prediction about how its interactions with the environment will affect the inputs that it receives. Thus, in addition to inferring how the world is, the system also infers action policies that it expects will be best at minimising prediction error over long time scales. These action policies can then generate real-world behaviours in a way similar to ideomotor accounts of motor control, such as the one proposed by Wolpert and Flanagan (2001; cf. Wiese 2017, though see Pickering and Clark 2014, as well as Adams et al. 2013, for important differences between forward- and generative-model based accounts of motor control). In other words, the system can actively intervene on its environment in ways that aim to minimize the ‘surprising’

signals. This can either be done by a selective sampling of the surroundings aimed at obtaining more data (e.g. eye saccades), or by manipulating the environment in a way that minimizes signal uncertainty and brings the surroundings into a tighter fit with the current model (cf. Friston et al. 2009). The latter offers a very general mechanism for action control, as any behaviour of the system can be reconceived as aiming to minimize prediction error over the long term, and is therefore a crucial ingredient in the PEM framework's claim to unify perception and action under one model.

2.4 Precision optimization

One problem faced by a PEM system is that not all prediction errors will be the result of inaccurate predictions about the world, but rather some may simply be caused by volatility in the environment or noise in the system's sensory channels. Thus, a PEM system needs to account for uncertainty in the error signals it receives, and aim to minimize only those that constitute a reliable learning signal. PEM proposes to alleviate this problem by estimating the *expected precision* of the prediction errors (usually formalized as the inverse of their variance), which is used to weight different error signals according to their reliability. The precisions of prediction errors are context dependent, i.e. the precision that can be expected from different sensory modalities and in different environmental conditions will vary, and therefore need to be learned over time. This suggests the need for an additional, domain general control mechanism, either sitting on top of the main prediction/error hierarchy, or incorporated into the system and acting as a 'virtual' controller.³ Either way, the system needs "to represent the known unknowns" (Hohwy 2012, p. 5) which determine the precision of the incoming error signals.

Recently it has been proposed that the role played by the optimization of expected precision in the PEM framework is equivalent to that of attention (Feldman and Friston 2010; Hohwy 2012; Brown and Friston 2013). According to this proposal, *exogenous* attention is present in the case of "stimuli with large spatial contrast and/or temporal contrast (abrupt onset)" (Hohwy 2012, p. 6), and is a result of the system's expectation that strong error signals, i.e. ones standing out from the predicted noise level, will be more precise. *Endogenous* attention, on the other hand, is the result of a top-down process in which predictions about incoming stimuli increase signal gain in the relevant receptive fields. The function of endogenous attention is to decrease the uncertainty of future stimuli rather than reacting in response to unexpectedly strong incoming error signals (Hohwy 2012, p. 6). This type of attention can be illustrated using the example of visual search tasks, where the search is driven by a highly precise prediction of a particular stimulus that the subject needs to locate. According to PEM, this increases

³ Engineers speak of a 'virtual' controller (or governor), when a complex system behaves as if it was controlled by a separate governing system, but such behavior is an emergent property of the whole system with no discrete parts playing the function of a governing system. Hooker, who introduced this term to the philosophical literature, explains that "'has a virtual governor of reliability f' is a property of the system as a whole but of none of its components. [And...] from an ontological point of view there is nothing to the system over and above its components and the physical relations between them (however physically realised)" (Hooker 1981, p. 509). We think it is plausible that the mechanism responsible for precision estimation in the PEM framework might be a virtual governor of this kind (cf. Dolega and Dewhurst 2019).

the gain on the receptive field sensitive to the object of the search. Once a pattern of retinal activations similar to the target is found, the high gain on the error signal pathways guarantees that only a hypothesis tailored to minimizing that very precise signal (corresponding to the object of the search) will be selected for driving the system's behavior. This precision/attention mechanism will play an important role in our deflationary account of consciousness in the PEM model, and we will return to it later (in Sect. 4).

2.5 Conscious experience as the 'winning hypothesis'

Finally, we arrive at the most popular proposal for how to explain consciousness within the PEM framework, namely that “conscious perception is determined by the prediction or hypothesis with the highest overall posterior probability—which is overall best at minimizing prediction error” (Hohwy 2012, p. 4).⁴ To put it simply, what we perceive is the brain's best guess about ‘what is out there’. This simple proposal is expanded upon by Hohwy (2012), who points out that by assuming this perspective, PEM ends up treating attention and perception as “two distinct, yet related aspects of the same prediction error minimization mechanism”, that together are responsible for determining “which contents are selected for conscious presentation” (Hohwy 2012, p. 1). At the core of this approach are two properties that control the dynamics of the bidirectional (i.e. top-down and bottom-up) exchange of information within the PEM system—accuracy and (expected) precision (Hohwy 2012, p. 5)—which together form the “statistical dimensions of conscious perception” (Hohwy 2012, p. 5), a state space within which conscious and unconscious perceptual states can be located and compared (see Fig. 1).

Hypotheses with high accuracy can be understood as being better at representing the causal structure of the world, since accuracy can be taken to refer “to the inverse amplitude of prediction error per se” (Hohwy 2012, p. 5), representing the degree to which a hypothesis can account for the incoming sensory signal. However, as will become relevant in the later parts of this paper, Hohwy points out that this characterization involves a slight simplification, as “surprise has both accuracy and complexity components, such that minimizing surprise or free energy increases accuracy while minimizing complexity” (Hohwy 2012: ft. 2). What this means is that accurate hypotheses are not only good at minimizing prediction error, but can accomplish that by recruiting fewer resources (e.g. by making use of models with fewer parameters). As Hohwy explains: “This ensures the explanations for sensory input are parsimonious and will generalize to new situations [...]” (Hohwy 2012: ft. 2).

Precise hypotheses, on the other hand, are those that are best at minimizing prediction error relative to the system's prior expectations of the error signals' variability. Thus, precise hypotheses are those that are best at explaining the error that is least random for the given prior distribution, i.e. they are best at ignoring noise and picking out

⁴ Hobson and Friston (2014) have made an alternative proposal with a slightly different flavour, but as Dołęga and Dewhurst (2015) argue, this approach strays too far into the realms of the ‘Cartesian materialism’ that Dennett (1991) warns of. Hobson and Friston have acknowledged this criticism of their proposal in a subsequent publication (Hobson and Friston 2016), suggesting a revised approach that takes on board some of these concerns.

Fig. 1 A diagram depicting the relationship between conscious and unconscious states with their accuracy (x-axis) and precision (y-axis) (from Hohwy 2013, p. 203)

the input that is relevant for a given context (Hohwy 2012, p. 4). Choosing the ‘best’ hypothesis involves making trade-offs between these two dimensions, as sometimes the most precise hypothesis might not be the one that is most accurate, and vice versa.

Although this proposal is centered around the basic assumption that conscious perception will typically belong within the space of high accuracy and precision (Fig. 1), Hohwy admits that this might not always be the case. It is possible that “a relatively inaccurate but precise model might determine conscious perception over a competing accurate but imprecise model, and vice versa” (Hohwy 2012, p. 5). The latter is what supposedly happens in gist perception, where subjects can perceive the general gist of a briefly presented scene, even though they are unaware of the details (see Bayne 2016 for a comprehensive overview). Conversely, the ‘ventriloquist effect’ (Jack and Thurlow 1973), involving a misperception of the location from which a sound comes, is an example in which our perceptual discrimination is highly precise and yet still inaccurate.

The role of attention in bringing predictions into consciousness is, therefore, twofold. Firstly, attention can raise a perceptual hypothesis into consciousness by adjusting precision weightings on either a strong incoming error signal (exogenous attention), or a signal expected to be highly reliable/salient in a given context (endogenous attention). Secondly, it can also serve as a kind of ‘bias control’ in cases of competition between hypotheses (Hohwy 2012, p. 7). Hohwy links this second role of attention to studies on the receptive fields of retinotopic neurons, which indicate that a single neuron’s activity is decreased “when two different objects are present in the neuron’s receptive field” (Hohwy 2012, p. 7). Following Feldman and Friston (2010), he proposes attention as the mechanism responsible for biasing neuronal responses between different inputs falling within the same receptive field, thus resolving the conflict and preventing further decrease of neuronal activity in that field. This role of attention is crucial for the PEM framework, as only one hypothesis can drive the system’s behavior in a particular task at any given time. This is consistent not only with the empirical findings within receptive fields studies (e.g. Robert Desimone’s famous research of biased competition, see Desimone and Duncan 1995; Desimone 1998), but

also with research on perceptual phenomena like binocular rivalry (Dijkstra et al. 2016; O'Shea et al. 2013), in which neurotypical subjects experience alteration between percepts that are simultaneously presented in each of their visual hemifields. In the case of the latter phenomenon, PEM conceptualizes the two percepts (e.g. a house and a face) as distinct hypotheses generated by the system in order to try and predict the current state of the visual environment. The perceptual system is unable to resolve the rivalry between these competing hypotheses because both are equally probable given the sensory evidence, nor can it synthesize the two into a singular percept due to their incompatibility with prior expectations about the structure of the environment (i.e. that objects belonging to separate categories cannot occupy the same spatiotemporal location). This is why the competition is resolved sequentially over time, resulting in an alternating percept. Although the heuristic description provided here is quite simple, a computational model inspired by this approach has recently been validated with fMRI data collected by Weinhhammer et al. (2017), who showed that the PEM model outperformed other competing accounts in predicting increased responses from cortical areas implicated in controlling the rate of perceptual bistability.

To sum up, Hohwy's proposal is that the status of a hypothesis as conscious or unconscious is determined by two separate dimensions: precision and accuracy (see Fig. 1). A hypothesis that is both precise and accurate will play a strong role in determining consciousness, whereas a hypothesis that is lacking in both dimensions will not, with other hypotheses falling somewhere along a spectrum from conscious to unconscious. One advantage of this proposal is that it can account for some variation in the quality of different conscious states: "inattentive but conscious states would cluster towards the lower right corner and attentive but unconscious states would cluster towards the upper left" (Hohwy 2012, p. 203). The proposal is also consistent with some recent findings that have been used to develop the PEM framework (such as the binocular rivalry studies described above), thus lending it a degree of empirical credibility.

3 Two problems with the 'winning hypothesis' approach

Our aim in this section is to explore the tacit assumptions behind the 'winning hypothesis' approach, revealing several conceptual problems that the framework is still facing. We saw earlier that the PEM framework assumes a direct connection between the content and properties of the generative model and the content and properties of conscious experience. It is relatively uncontroversial to assume that the content of perceptual experiences is a product of subpersonal processes, and that cognitive functions like attention are involved in determining which contents are made available for conscious presentation. However, the way in which numerous perceptual paradigms are framed in PEM explanations (see e.g. Clark 2013, 2016; Hohwy 2012 for an overview) implies that the framework carries a stronger implicit assumption about the relationship between the winning hypothesis (or state of the internal model) and the

phenomenal aspects of conscious experience. Namely, most proponents of PEM seem to be committed to a form of representationalism about conscious experience.⁵

Although there are many varieties of this philosophical position (see Chalmers 2004 and Lycan 2019 for comprehensive overviews), in its basic form representationalism about conscious experience is the thesis that the phenomenal aspect or properties of conscious experience are part of, or coincide with, the content which is represented by that experience.⁶ Proponents of strong versions of the view (e.g. Dretske 1995; Tye 1995) postulate that representation of a certain kind is *sufficient* for phenomenal properties (“where the kind can be specified in functionalist or other familiar materialist terms”, Lycan 2019, p. 2.1) or that such properties *reduce* to representational ones (Kind 2007, p. 406). Weak representationalists (Block 1990, 1996; Chalmers 1996), on the other hand, are only committed to the view that phenomenal properties (of a certain kind or within a certain class) *entail* representational properties (of the same kind or class), but not vice versa, meaning that some further story is required about how phenomenal properties are determined.

A commitment to some form of this view is clear from the deployment by PEM proponents’ of the argument from transparency, which has long played an important role in the representationalists’ argumentative arsenal (cf. e.g. Moore 1903; Harman 1990; Tye 1995, 2003; Dretske 2003; Metzinger 2003, 2008). According to this argument, our perceptual experiences are transparent or diaphanous to us because our introspection into ‘what it is like’ to experience something does not reveal that we are in a perceptual state, while “the properties we are aware of in perception are attributed to the objects perceived” (Lycan 2019, Sect. 4.3). For example, when I see a red tomato, I experience the redness of the tomato as a property of the object which is the content of my conscious state, and not as a property of that conscious state itself. On the PEM side, Limanowski and Friston (2018) propose that this property of our experiential states “can metaphorically be understood as looking through a window onto the world, instead of looking at the window itself: we only access the representation’s intentional content (something in the world which it is about) without noticing its non-intentional carrier properties” (2018, p. 4), revealing a clear commitment to representational transparency.

Hohwy (2014, pp. 184–185) also embraces representationalism by arguing that PEM satisfies key requirements for a successful theory of phenomenal experience as laid out by Frank Jackson (2003). According to him, conscious representations in PEM are not only rich in detail due to an underlying “variance of percepts from

⁵ Note that representationalism about conscious experience is a distinct issue to the representational status of the subpersonal mechanisms that constitute the PEM system. Although most proponents of PEM are also representationalists in this latter sense, there have been some dissenting voices, and we will not address that issue here (see Dolega, 2017 for an overview/discussion of the representational status of PEM).

⁶ A further distinction among different representational views which has been made in philosophy involves that between viewing experiences as carrying Russellian or Fregean contents. In the case of the former, the content “can be a singular proposition or a configuration of objects and their properties.” (Lycan 2019: 2.3). Although such contents can be presented under a certain description or mode, the manner in which they are presented is not taken to be a part of the content itself. Fregean contents, on the other hand, do include the mode of presentation as part of the content of representations, but exclude objects from being part of the content (cf. Lycan 2019: 2.3). The Russellian position has been famously defended by Tye (1995, 2003, 2009) and Dretske (1993, 1996), while Chalmers (2004) has defended its Fregean counterpart.

the first-person perspective as represented in the fast time-scale causal regularities in the lower levels of the perceptual hierarchy” (Hohwy 2014, p. 184), but can even be said to be “inextricably rich” due to the “richly hierarchically structured perceptual inference (including binding, penetrability, and sensory integration); [as on PEM] there is no perceptual inference without a perceptual hierarchy to accommodate empirical Bayes.” (ibid.).⁷

Yet, despite this, it is rather unclear which version of the view (strong or weak) Hohwy subscribes to. On one hand he states that PEM “seems to tick the boxes that Jackson and others have set out as requisites for a substantial representationalism about consciousness.” (Hohwy 2014, p. 184). While it is open to interpretation whether ‘substantial’ is meant as synonymous with ‘strong’, it seems quite clear that a strongly-representational position would be the best fit for PEM’s explanatory ambitions, as it would allow for an explanation of consciousness without relying on resources external to the framework. However, Hohwy undercuts this interpretation by also claiming that PEM is unlikely to solve the mystery of phenomenal experience and that the best it can hope for is to offer some insight into the nature of the mechanisms that bring about key aspects conscious perception (such as perceptual binding, illusions, and mental illness) and to help us capture “many of the things we care about in conscious experience” (Hohwy 2014, p. 208). Such statements side with a weakly representational view, but are inconsistent with his commitment to the framework’s explanatory power and ambitions.

In the remainder of this section we will investigate two problems emerging from these competing interpretations of PEM’s representational commitments, and show that either interpretation carries a significant cost for its proponents.

3.1 The problem of unconscious representation

Let us now return to the previous example of binocular rivalry. According to PEM, the subject might have the experience of a face, rather than a house, because the ‘face’ hypothesis is currently the winning one and the ‘house’ hypothesis is suppressed (Hohwy et al. 2008). Therefore the face hypothesis determines the content and phenomenal quality of the conscious state (at this time). However, the house hypothesis is also represented in the system at the same time. The default view about how hypotheses come to determine conscious experience, when paired with PEM’s commitment to representationalism about consciousness, seems to lead to a puzzle.

If PEM is committed to the strong view and phenomenal properties are identified with representational properties, then proponents of the framework need to spell out why the competing hypotheses are not both experienced phenomenally at the same time. After all, what changes in the process of switching are the probabilities assigned to such hypotheses, not their contents. A common answer to this kind of puzzle is that strong representationalism requires representations of *a certain kind* for content to be conscious (see e.g. Tye 1995). This may be helpful to PEM proponents in the short run, but eventually they will have to face a further question: what is it about the attribution of highest posterior probability that makes contents conscious?

⁷ An assumption that we will, to a certain extent, challenge in Sect. 4.

Perhaps it is the difficulty of answering this second question that pushes Hohwy to sometimes interpret PEM as being committed only to weak representationalism (see e.g. Hohwy 2014, pp. 208–209). Moving in this direction helps to sidestep the question altogether, since weak representationalism does not bind one to the claim that phenomenal properties *are* representational properties. However, it forces a proponent of the framework into a sort of quietism in which they have to state that, even though representations with the highest posterior probability do determine conscious perception, PEM does not answer why or how such states contribute to the phenomenal aspect of experience.

Although Hohwy sometimes seems to side with the latter option, he also acknowledges that PEM does face the problem of answering whether representations that are suppressed, but nevertheless processed by the system, do come with phenomenal properties:

There remains the rather important and difficult question whether or not the unseen stimulus is in fact consciously perceived but not accessible for introspective report, or whether it is not consciously perceived at all; this question relates to the influential distinction between access consciousness and phenomenal consciousness (Block 1995, 2008). (Hohwy 2012, p. 7)

Here, he refers to the distinction between the phenomenal, qualitative aspect of a state and the availability of that state to introspection, report, and other forms of executive functions, i.e. its accessibility to consciousness. Ned Block (1995, 2008) and Victor Lamme (2006) have argued that all that is needed for a state to be imbued with phenomenal quality is that it is sustained over time by recurrent activity within a sensory cortex, even in the absence of any kind of cognitive access to that state (Block 2007). Howhy does not directly comment on this issue, but he does seem to side with Block and Lamme when he speculates that:

(i) access consciousness goes with active inference (i.e. minimizing surprise though agency, which requires making model parameters and states available to control systems), and (ii) phenomenal consciousness goes with perceptual inference (i.e. minimizing the bound on surprise by more passively updating model parameters and states). (Hohwy 2012, p. 7)

While the above proposal may solve the present problem (by effectively embracing the view that hypotheses that are represented in the PEM system but are inaccessible to the subject do come with phenomenal qualities), it ends up introducing additional confusion about the position held by Hohwy. Firstly, the proposal is ambiguous between embracing an unusually strong version of representationalism on which all that is needed for content to instantiate phenomenal properties is that it is rendered in the system via perceptual inference, and claiming that perceptual inference is somehow instrumental, but not sufficient, for the instantiation of phenomenal properties. Embracing the former position leads to a proposal which may not be coherent—how can one be in a phenomenally conscious state, and yet not be aware of it? How could I be in a qualitative state that ‘it is something like to experience’, yet not know ‘what that state is like’? Cohen and Dennett (2011) have criticized this position and pointed out that the very possibility of phenomenal experience in the absence of cognitive access

is not verifiable in any way, since all empirical methods of investigating consciousness require the availability of some form of explicit or implicit report. Embracing the latter option—that only perceptual and not active inference plays a role in instantiating phenomenal properties—is less problematic, but it is significantly underdeveloped, and it is unclear whether such an interpretation is consistent with Block’s original phenomenal/access distinction (as Block and his followers associate phenomenal consciousness with particular *stages* of perceptual processing rather than *kinds* of processing). However, the most unwelcome consequence of embracing either solution is that neither of them can help us understand how the initial proposal that consciousness is determined by the hypothesis with the highest posterior probability can account for the well documented phenomenon of unconscious perception, to which we turn next.

3.2 The problem of unconscious perception

Our ability to respond above chance to stimuli that we are not consciously aware of is well documented in the literature. It seems that some perceptual signals are simply too weak to ever become conscious, but can nonetheless influence our behaviour. For example, subjects fail to detect subliminal primes even in controlled settings, but such stimuli can still drive behavioral responses (see e.g. Norman et al. 2014; Faivre et al. 2014; see also Ansorge et al. 2014 for a review). On the other hand, several studies have shown that percepts which were not consciously registered by the subjects can later be brought into consciousness by attentional cues directed at the location of the stimulus after its disappearance (Kentridge 2013; Sergent et al. 2013; Thibault et al. 2016). This is important, as it lends support to the PEM idea that, by modulating the probability of the bottom-up signals, attention determines how much influence their probability should have on the posterior. However, this also seems to pose a problem as it indicates that whether or not some content is available for conscious perception is not just a question of its relative probability, but also other factors (such as subsequent attentional cues).

Consider the cases of blindsight and visual form agnosia. The former occurs when patients who have suffered lesions to the primary visual cortex become ‘blind’ on one side of their visual field, but nonetheless retain some discriminatory ability on that side (see e.g. Humphrey 1974; Weiskrantz 1986; Holt 2003). The latter is a condition in which localized damage to the temporal lobes causes loss of the capacity to recognize shapes or objects, but leaves vision-guided motor coordination mostly unaffected (see e.g. Benson and Greenberg 1969; Milner et al. 1991; Carey et al. 1996). Patients suffering from these conditions have been shown to detect, localize, and act upon visual stimuli apparently without experiencing them consciously, which in turn has been proposed as evidence that perception can affect behavior in the absence of consciousness. For example, one particular subject suffering from visual form agnosia, known as D.F., has been shown to retain the ability to perform simple visuo-motor tasks (like grabbing a cup or posting a letter into a letterbox) nearly or even equally well as healthy controls, despite not being able to discriminate the shapes of objects involved in the tasks (i.e. performing at chance in shape or object discrimination tasks).

When looking at such cases from the perspective of PEM, it seems that the visual system does form distinct, albeit imprecise, hypotheses about the objects present in the portion of the visual field affected by the cortical damage. Since such hypotheses are able to affect behavioral responses, they must at some point be assigned the highest posterior probability, i.e. be the ‘winning hypothesis’ on some level of the hierarchy. Yet they are not experienced in consciousness, and so posterior probability alone cannot be the only property that determines whether a hypothesis is selected for conscious presentation.

Squaring such evidence with Hohwy’s earlier stipulation that phenomenal consciousness is associated with perceptual inference is also difficult. It seems that, by subscribing to Block’s distinction, Hohwy’s original proposal that consciousness is determined by the hypothesis with the highest posterior probability becomes severely limited in scope. Embracing Block’s distinction means that the winning hypothesis view is only an account of access consciousness and not phenomenal consciousness. However, even that might not work once we consider visual form agnosia, which implies that sometimes unconscious states *do* take control of the system’s behavioral responses, in the same way that conscious states can, but without any (reported) phenomenal experience. Effectively this means that, at least in some cases, hypotheses can be ascribed the highest posterior probability while remaining unconscious, which in turn undermines the idea that the ‘winning hypothesis’ view offers a complete account of the functional aspect of consciousness.⁸

Hohwy anticipates this problem by pointing out that the ‘winning hypothesis’ view does not necessarily apply to hypotheses on a single level of the hierarchy, but rather should be understood as targeting hypotheses that successfully minimize prediction error across multiple levels:

When a specific model is the one determining the consciously perceived content it is just because it best minimizes prediction error across most levels of the cortical hierarchy—it best represents the world given all the evidence and the widest possible context. This is the model that should be used to selectively sample the world to minimize surprise in active inference. Competing but less probable models cannot simultaneously determine the target of active inference: the models would be at cross-purposes such that the system would predict more surprise than if it relies on one model alone (for more on the relation between attention and action, see Wu 2011). (Hohwy 2012)

According to Hohwy, then, there is no specific threshold for consciousness, but his answer to the question implies that whether a hypothesis is accessible depends on the landscape of probabilities in which it competes. The reason why a relatively inaccurate and imprecise hypothesis can sometimes be the one driving conscious experience

⁸ Some may object to the idea that unconscious states that drive overt behavior are functionally equivalent to conscious states by pointing out that the former are not available for verbal report or introspection. There are two problems with this response. Firstly, it is question begging, as there needs to be *some* observable difference between such states for us to be able to discover them in the first place. Secondly, the performance of visual form agnosia patients on motor tasks implies that the inaccessibility of states driving their performance is likely not caused by changes intrinsic to those states, since any impairment to their functional profile would presumably have a negative effect on task performance.

is because all the competing models are ‘worse’. The winning hypothesis, simply put, merely needs to be better than the competition. This move pushes back on our accusation that the ‘winning hypothesis’ cannot account for the functional aspect of conscious experience, while embracing the proposal that it just *is* an account of access consciousness.

All of the above seems to support the weakly representational interpretation offered at the end of Sect. 3.1. While this may be a satisfying conclusion for many, and the position certainly has the virtue of being restrained, we would like to (once again) stress that, at least at face value, it is inconsistent with PEM’s ambition to explain all facets of our mental lives (Hohwy 2013, p. 1). Embracing such a position, therefore, is an admittance of PEM’s limited explanatory power and inability to deliver on its promises. In the next section we will argue that this view be strengthened by rejecting Block’s access/phenomenal distinction, which would enable it to offer an answer to questions about the nature of phenomenal consciousness (albeit a deflationary one), and therefore rescue some of PEM’s broad explanatory ambition.

4 A deflationary approach to PEM and consciousness

Having presented an overview of PEM and assessed one popular proposal for how it might account for consciousness, we will now turn to our own alternative position. Our approach has much in common with Hohwy’s position, insofar as it retains an important role for attention in determining conscious content, but it moves away from the idea that high probability alone is sufficient for content to be conscious.⁹ Our proposal is that in order to (potentially) enter consciousness, a hypothesis must be ‘probe-able’, a feature that depends partly (but not entirely) on its probability. As we will explain in more detail shortly, the Dennettian concept of a ‘probe’ that determines consciousness fits nicely with the concept of active inference and the role of attention in PEM, and so we will propose that the PEM mechanism responsible for controlling attention can serve as an implementation of probing. By adopting this approach, we will reframe Hohwy’s original proposal as a self-standing approach to consciousness and resolve the problems it faces.

Before presenting our proposal in more detail, we will first outline some important philosophical assumptions that underwrite this approach. Our aim here is not to present a defense of these assumptions, but rather to clearly state where we are coming from, so that any disagreement will (hopefully) be made more perspicuous. We will then explain how we think this approach helps us to resolve (or dissolve) the two problems outlined above, before moving on in the final section to consider some further outstanding issues.

⁹ An anonymous reviewer has pointed out that, given what has been said about Hohwy’s position in Sect. 2 and 2.1, it is possible that he might be defending a much weaker position that allows for fringe cases in which the content of conscious experiences is to be explained not in terms of high probability, but rather some other (as yet unspecified) factors. While we admit that this is a salient possibility, we haven’t found any evidence that this is the position embraced by Hohwy, or any other proponents of the framework. We would also like to point out that subscribing to such a position would move Hohwy’s view even further away from the initial goal of PEM being a self-sufficient framework for explaining cognition.

4.1 The method of heterophenomenology

There is widespread agreement that any serious empirical investigation into conscious experience must take subjective verbal reports as a genuine datum (see e.g. Piccinini 2003; Chalmers 2013). To explain consciousness we need to explain why we experience and report our experiences as being this way rather than another (Chalmers 2018). However, this basic intuition clashes with the fact that first person reports about subjective experience, while a valuable source of data about perceptual processing, are often considered to be unreliable due to the lack of any possibility for verification (Dennett 1991; Schwitzgebel 2007, 2012b). Medical conditions like visual anosognosia, in which subjects are unaware that they are blind (Prigatano and Schacter 1991); visual illusions, like the color phi phenomenon (Kolars and von Grünau 1976); and the fact that subjects systematically overestimate the precision and richness of visual information outside of the foveal area (Dennett 1991) and outside of the scope of attention (Knotts et al. 2018) all indicate the general unreliability of our introspective abilities (see e.g. Schwitzgebel 2007, 2012a, b for further discussion).

Dennett's 'heterophenomenological' methodology (1991, 2003, 2006) is aimed at taking both of these points seriously. In contrast to the phenomenological tradition (see Smith 2018) and introspection based psychology (e.g. Titchener 1901), heterophenomenology does not consider subjective self-reports as ultimately authoritative about mental phenomena, but rather treats them as non-committal fictions, which carry information about how things seem or are judged to be by the subjects, rather than how they really are. In this regard heterophenomenology is claimed to be similar to anthropology, as it suspends judgement about the reality of the entities and properties of subject's reports, in a similar way to how an anthropologist studying some peoples' religious system does not pass judgement on the existence of deities and other entities that are the object of those peoples' beliefs. As Dennett points out, this way of doing science of the mind "is nothing new; it is nothing other than the method that has been used by psychophysicists, cognitive psychologists, clinical neuropsychologists, and just about everybody who has ever purported to study human consciousness in a serious, scientific way" (2003, p. 22).

Associated with this proposal is the rejection of the idea of phenomenal properties (or 'qualia') as ineffable and intrinsic properties of experiential states, to which conscious subjects have immediate (and privileged) epistemic access (cf. Dennett 1988). Since properties present in experience can be in some sense illusory (e.g. peripheral vision appears as rich and detailed even though it is not, colors appear to be the intrinsic properties of objects even though they are bundles of complex relational properties; cf. Cohen 2009), consciousness researchers should not aim to explain the existence of such properties themselves, but should rather explain why conscious experience appears to have such properties (Frankish 2016a, b, cf. Chalmers 2018), i.e. why and how subjects come to believe that their experiences have such properties.

As mentioned before, our approach in the remainder of this paper will be fundamentally deflationary, in just the way implied by heterophenomenology. We take the strong representational proposal that phenomenal properties are identical to representational properties quite seriously, and the solution to the problems posed in the

previous section will simply require an explanation for why some hypothesis (instead of another) determines the content of a verbal report about conscious experience, rather than determining the phenomenal properties of that experience as such. Following Dennett, we think that taking this course of action will result in rendering the problematic or mysterious character of phenomenal properties as nothing more than an adaptive illusion. Of course, in line with the outlined method, and in order to fully amend Hohwy's account, we will also need to say why or how such an illusion comes about. With these preliminary remarks out of the way, we will now move on to our main positive proposal.

4.2 Multiple-drafts model and the 'fame in the brain' metaphor

The PEM account of consciousness developed by Hohwy (and others) bears some interesting similarities to Dennett's multiple drafts (or 'fame in the brain') approach, which we think positions it well to adopt the further innovation of a 'probe' that determines conscious content. According to Dennett's model of consciousness there are multiple, parallel streams of information that are being processed in the brain at any given time. Conscious experience is determined by a 'probe' that, through a process of eliciting reports (verbal or otherwise) from a subject, takes one of the competing information streams as a more-less complete story about what is going on in the mind at that time (1991, p. 135). Probing may be verbal, environmental, or self-induced; what matters is that the subject is able to respond to the probe in an appropriate way, i.e. by producing an internal or external indication of their apparent conscious experience (e.g. a silent or vocalized report, or the production of some behavior that is consistent with the apparent conscious experience). In this context a 'probe' is a kind of higher-order, meta-cognitive state which takes another perceptual or cognitive state as its object. Prior to being probed in this way, there is no determinate fact of the matter about what the subject is conscious of, but rather a stream of possible 'drafts', each of which is potentially available to determine the content of a report if it is probed in the right way (we could say that a draft of this kind is 'probe-able'). As Dennett puts it, there are "multiple 'drafts' of narrative fragments at various stages of editing in various places in the brain. Probing this stream at different places and times produces different effects, precipitates different narratives from the subject" (Dennett 1991, p. 113).

Although it is impossible to pinpoint which contents will be selected for consciousness prior to a report, any process can have more or less 'fame' or 'clout' (Dennett 2005, pp. 136–138), i.e. be more or less likely to feature in a subject's report, in virtue of the functional role it plays in global information processing in the brain. Thus, the question of which states and processes 'become conscious' boils down to a distinct question of which of many possible 'drafts' (i.e. information streams) is the one exerting the most influence on the system at the time of the probe, and thus which draft becomes available for report (verbal or otherwise). While it may seem counterintuitive, this means that, on Dennett's conception, there is actually no fact of the matter about what a system is conscious of until it is probed (although it is important to note that a probe may be either internally or externally generated, meaning that a system might become conscious 'spontaneously', at least as it appears from the outside). Dennett's

metaphor of ‘fame in the brain’ is meant to refer to how likely a state is to be in control of the system at the time it is probed, and thus how likely it is to determine the content of consciousness at that moment.

While we do not wish to subscribe to the claim that the contents of unconscious states are underdetermined prior to probing (according to PEM, there is a fact of the matter about which hypothesis is currently ascribed the highest posterior probability, even if its contents are inaccessible to the subject), we do think that the notion of ‘fame in the brain’ is extremely similar to the idea of the winning hypothesis being responsible for determining the contents of conscious experience (described in Sect. 2.5), while the notion of probing can be treated as equivalent to the way that PEM describes attention and active-inference (Sects. 2.3 and 2.4). We develop this proposal in more detail below.

4.3 Probe-ability determines conscious content

Probing, we want to suggest, can be cashed out in terms of the PEM notion of attention. A probe can be considered a higher-order process of predicting the states of the system in accordance with the expected precision of the incoming error signals.¹⁰ As we have already suggested in Sect. 2.4, for attention to be successfully deployed the system needs to draw on stored information about the context in which the prediction errors are generated. In other words, it needs to draw on the multi-level representations which shaped the hypotheses that generated the erroneous predictions, in order to assess the reliability of the errors related to said predictions. Thus, the successful deployment and control of attention on a given level of the hierarchy will require the system to issue descending predictions about the expected states on that level of the hierarchy, in order to predict what precision the errors on a given level should have. Effectively, the system needs to be predicting its own states in order to deploy attention, and these top-down meta-predictions, we want to suggest, constitute what Dennett calls a probe.¹¹

¹⁰ One important kind of probe that we unfortunately do not have space to discuss in detail here are *linguistic* probes, i.e. self- or other-directed natural language that elicits a conscious report (verbal or otherwise). As language has also been suggested to play a crucial role in attention modulation according to some versions of the PEM framework (see e.g. Lupyan and Clark 2015), it seems like there could be an interesting avenue of research here, one that we would be keen to explore in future work. We thank an anonymous reviewer for bringing this point to our attention.

¹¹ We do not want to associate probing with the activity of a particular model within the hierarchy, but rather suggest that (as discussed in 2.4) it can be treated as the output of a distributed ‘virtual governor’. This means that we are safe from the potential objection that we are postulating a ‘Cartesian theatre’ (Dennett 1991) and do not fall into the same issues as, e.g. Hobson and Friston (2014; cf. Dołęga and Dewhurst 2015). It does, however, raise an additional interesting question, of how to individuate instances of probing (we thank an anonymous reviewer for pressing us on this point). If *every* instance of a top-down meta-prediction constitutes a probe, then these are going to be much more ubiquitous than they seem to be in the original Dennettian story, perhaps resulting in something close to Graziano’s (2016) attention schema theory of consciousness. On the other hand, if only some of these meta-predictions constitute probes, then we will need to provide an additional story about which ones qualify as probes, and why. Our preliminary answer is that the meta-predictions that qualify as instances of probing will typically be the ones affecting the most levels in the hierarchy, often having a relatively gross behavioural outcome. It also seems plausible that the kind of pooling of contextual information that is required for generating predictions of precision

Recall that, according to PEM, attention is the allocation of expected precision based on the relative spread of the sensory samples. There are two distinct ways in which attention can influence the attenuation of prediction error. Exogenous attention operates by raising the precision of predictions that can account for sudden or especially pronounced error signals, increasing the strength of the hypothesis that generates those predictions and thus making it more likely that that hypothesis will come to dominate the hierarchy. This can be likened to the way in which a sudden, unexpected event can act as a probe, generating a conscious report from a system that was previously operating ‘on autopilot’, so to speak. We can illustrate this with the example of driving down a familiar street, not really paying any (conscious) attention to one’s surroundings, when another car pulls out in front of you and forces you to suddenly slam on the brakes. Even though prior to this event there may not have been any fact of the matter about what one was ‘conscious’ of, immediately afterwards you will have a seemingly conscious memory of what was happening prior to the car pulling out (cf. Dennett 1991, 137). From the PEM perspective, this seems to be a clear case of exogenous attention determining the content of conscious experience by raising previously unconscious hypotheses into awareness, in a way that fits very neatly with the Dennettian story about probing.

Endogenous attention, on the other hand, operates in a top-down fashion, where the system raises the expected precision of predictions matching the object of attention, thereby increasing the salience of sensory inputs that confirm these predictions. This type of attention can be illustrated using the example of *Where’s Waldo?* books. The premise of these books is that the reader has to find the character Waldo, who is hidden in complex, confusing images filled with human and animal figures. The search is driven by a top-down prediction of a particular stimulus, Waldo, appearing against the noisy background. Thus, this process relies on the mechanism of endogenous attention—the gain on the receptive field sensitive to the object of the search is increased. Notice that this requires the attentional control system to form a high-level hypothesis about the target of the search. Once a pattern of retinal activations corresponding to those possibly caused by the target is found, the high gain on the error signal pathways guarantees that only a hypothesis tailored to minimize that very precise signal (congruent with the ‘Waldo hypothesis’) will be selected for driving the system’s behavior. In this case, attentional modulation comes from within the system itself, highlighting the draft (a hypothesis or a collection of hypotheses on multiple levels) that corresponds to the behavior of scanning the page looking for Waldo. Thus, the system will be primed to report a conscious experience with this content, though importantly other ‘shorter’ drafts can also capture a system’s resources in an exogenous manner during visual scanning.¹²

Footnote 11 continued

across multiple levels of the hierarchy may require long distance signaling between different cortical areas, the kind of which are postulated by the global workspace hypothesis (Baars 1997; Dehaene and Naccache 2001). Thus, probes as such may be distinguishable from meta-predictions in general by their distinct functional and neural profiles. We hope that this might preserve the Dennettian spirit of our proposal while accommodating the details of the PEM account of attention.

¹² One of the interesting consequences of the PP account of endogenous attention is that the system needs to de-regulate lower level perceptual states in order to boost the error signal on the pathways relevant for the search task. This means that, in the case of searching for the Waldo figure, the system needs to fix

Both kinds of attention contribute to determining the winning hypothesis and thus, according to Hohwy, the content of conscious experience. Here the concept of a ‘winning hypothesis’ can be likened to Dennett’s metaphor of ‘fame in the brain’—the hypothesis that currently enjoys the greatest control of the system is the one that, when probed, determines the content of experiential reports. Our proposal is that a winning hypothesis (or, in the case of a more complex stimulus/situation, a hierarchically organised bundle of hypotheses) becomes more probe-able the greater its probability. Note, however, that on this view it is not sufficient for a process to be highly probable for it to be conscious, rather the system needs to be actively probing the model—in other words, it needs to be performing active inference on its own states. It is this additional feature, coupled with the Dennettian deflationism about the phenomenal status of a draft (or hypothesis) pre-probe, that gives us the tools to resolve the two problems introduced in the previous section.

A possible solution to what we previously called the problem of unconscious perception should now be visible. Although phenomenal properties simply *are* representational properties, the difference between conscious and unconscious hypotheses lies in which of these states the system is currently conducting active inference on. This means that, in cases like binocular rivalry, there may be multiple candidate hypotheses, each of which is sufficiently high probability to be ‘probe-able’, but none of which will determine the content of consciousness until the moment that one is actually probed. The attention mechanism described above provides a simple account of how this process of probing can take place: attending to a hypothesis heightens the precision (and probability) of that hypothesis, boosting it to a level where it is able to (temporarily) take control of the system and thus elicit a verbal report (or some other indication of consciousness according to the heterophenomenological method). In the case of exogenous attention this is triggered by an external factor, such as a loud noise or sudden change in the visual field, whereas in the endogenous case it is triggered internally, and is the result of the system predicting how its lower-level internal states should be in order to minimize long term prediction error.¹³

This approach can also help explain why an antecedently ‘improbable’ hypothesis (low accuracy, low precision, or some combination of the two) can sometimes nonetheless enter consciousness: the very act of probing serves to focus attention on that hypothesis, heightening its precision and thus boosting its ‘fame’. A clear example of this kind of situation is when a sudden, surprising event focuses our attention on a previously unlikely hypothesis, such as a loud noise that we immediately become aware of. Similarly, an exogenous or endogenous probe can make salient a previ-

Footnote 12 continued

at least some portions of the hierarchy to states incongruent with the desired input for it to produce error signals to can be explained away by a higher-order ‘Waldo hypothesis’. Thus, the prior knowledge of what Waldo looks like biases the system to endogenously induce the pop-out effect when the desired stimulus is encountered. This kind of self-induced de-regulation may be the source of our introspective unreliability, as suggested by Dołęga (2018).

¹³ Dołęga and Dewhurst (2019) have described how this attention mechanism might also be integrated with Graziano’s attention schema model (2013, 2016), which we think can shed further light on the apparently mysterious nature of consciousness.

ously unconscious hypothesis, bringing ‘into’ consciousness the events of the last few moments, and making it seem as though we were always aware of them. Classic examples of this phenomenon include focusing on the last few chimes of a clock that one was previously ignoring, but nonetheless being able to accurately determine how many chimes there were (an endogenous probe), or being jolted into awareness by a surprising event while driving along a familiar route on ‘autopilot’, but nonetheless retaining a clear sense of having been aware of driving prior to the event (an exogenous probe). In both cases the Dennettian approach would argue that the question of whether or not one was conscious ‘all along’ simply doesn’t make sense, as prior to a probe there is just no fact of the matter about what one was conscious of, but rather a hierarchy of potentially probe-able drafts (or hypotheses) that are only selected from at the moment of probing. This approach, we think, fits very neatly into the existing PEM story about the determination of consciousness, helping to make sense of some otherwise puzzling features without having to posit any additional tools or mechanism (as hypotheses take the place of drafts, and probing is naturally identified with the deployment of the attention mechanism). It also sheds useful light on Dennett’s more recent work, in which he has defended attention as a resource which can be distributed among many perceptually salient targets at the same time (Cohen and Dennett 2011). This fits nicely with our proposal, as the conception of probing we described above allows for many ‘famous’ hypotheses located on different levels to be simultaneously probed by levels higher up the hierarchy.

Finally, the approach sketched here can give an account of unconscious perception without falling into the same issues as the winning hypothesis account. It is not surprising that patients suffering from partial loss of vision, as in the case of blindness, can perform certain visual tasks in the absence of consciousness. Such subjects do form hypotheses that are functionally capable of driving their responses on relevant tasks, however since most cases of unconscious perception are caused by serious pathologies resulting from structural damage, such hypotheses are likely inaccessible for probing and precision adjustment by the rest of the system. This may also explain why, despite performing adequately on relevant tasks, subjects suffering from such deficiencies have a very limited capacity for forming subjective assessments of their own reliability.

4.4 Cleaning up the phenomenal residue

So far we have applied Dennett’s multiple drafts model to PEM and shown that, on our view, what differentiates conscious from unconscious representations is not just their posterior probability, but also the fact that they are the target of a system actively generating meta-predictions about their optimal state and precision. Given our heterophenomenological assumptions, this is enough to deal with the problems incurred by unconscious perception and representation. However, critics of the deflationary approach may point out that all we have done is reframe Hohwy’s own view about active inference being linked to the winning hypothesis—an account of access consciousness—without really addressing the problem of phenomenal consciousness. While we do not hope to completely defuse this worry, we do think that our approach

can help to address at least two distinct questions that deflationary approaches sometimes face. The first question has to do with why certain worldly properties, e.g. electromagnetic waves of a certain length, are represented by contents with a distinct qualitative character, e.g. redness, while others, e.g. the number two, are not. The second question is about accommodating our intuitions that there is something more to subjective properties than them just being representations, i.e. that there is something ‘special’ or ‘mysterious’ about phenomenal consciousness, or what Chalmers (2018) has recently described as “the meta-problem of consciousness”.

The first question is largely an empirical one. For example, finding out why the human visual system is attuned to track certain physical quantities is largely a matter of exploring the morphological and ecological constraints that shaped our evolutionary history. However, the issue that typically troubles philosophers more is answering the second question: why do certain contents come with a specific subjective feel—i.e. why does that particular combination of physical properties cause such a sensation of redness, when it could have very well caused a sensation of greenness? Again, the exact nomological relation between, say, the space of possible physical quantities responsible for color perception and the space of phenomenal color sensations will likely depend on empirical facts about human vision. However, PEM can offer some insights into why certain clusters of quantitative physical properties might be represented by seemingly very different qualitative properties, i.e. why it is that quantitative properties can come to seem ‘special’ or ‘mysterious’.

Recall (from Sect. 2.5) that the successful minimization of prediction error involves minimizing surprise or free energy by increasing the accuracy of hypotheses, while simultaneously minimizing their complexity. What this means is that the PEM system is constantly making tradeoffs between speed and efficiency on the one hand, and devoting costly resources to fine-tune its representations of the world on the other. Faced with finite resources and external constraints, it is not surprising that a PEM brain might settle on furnishing internal models with only as much detail as is necessary for achieving a given task, compressing complex webs of interdependent physical causes into frugal representations that optimize not for veridicality, but rather for processing speed and efficiency. One result of this frugal representational format might be that our conscious states come to seem (to us) to be mysterious, as they are relatively lacking in detail (cf. Dołęga and Dewhurst 2019).

Combining PEM’s propensity to favor frugal representations with our proposal regarding probing can also shed further light on our intuitions about the mysterious or ineffable quality of consciousness.¹⁴ When we take into account the hierarchical nature of the PEM cognitive architecture, and the fact that probing is essentially a metacognitive capacity involving predicting the optimal state of the system’s models by drawing on prior information about the world and the context the agent is in, it turns to be relatively unsurprising that probing frugal models may lead to somewhat inaccurate inferences about one’s own capacities. The successful precision optimization of frugal perceptual models will require drawing on available contextual information about how such models operate, and using it to form predictions about how to optimize

¹⁴ Dewhurst and Dołęga (forthcoming) provide an explicit application of this approach to Chalmers’ meta-problem.

information processing. Thus, the successful deployment of endogenous attention, say to a particular perceptual process, will require forming a representation that is ‘twice-removed’ from the world—a facsimile of a facsimile of how the world is. Such a meta-representation will itself be optimized for complexity in the same way as its target representations are. This process of nesting content within content may be what is responsible for our intuition that there is more to the contents of our perception than just representational properties. However, this is just a ‘user illusion’ (Dennett 1991) caused by our cognitive system’s adaptive capacity for efficiency over accuracy. Thus, properties *of* and present *in* our experience only seem mysterious because of the way in which they come to be represented by the PEM system.

5 Conclusion

We have argued that Hohwy’s proposal for explaining consciousness within the PEM framework, while promising, suffers from two outstanding problems. Firstly, it seems to suggest the existence of phenomenal-yet-unconscious states, as there can be cases where two or more functionally equivalent states fulfil all the criteria for consciousness, but only one is experienced by the subject as being conscious. Secondly, and relatedly, it seems that Hohwy’s approach cannot adequately account for cases of unconscious perception, such as blindsight. We then suggested that both problems can be resolved by adopting a deflationary, Dennettian approach to the determination of consciousness, where multiple hypotheses can be simultaneously ‘probe-able’, but do not become conscious until they are actually probed, eliciting either a verbal or behavioural report. This approach is well-suited to the PEM framework, as both draw on similar bodies of empirical research for support (e.g. biased competition), and also share other structural similarities, such as the role played by attention (or probing) in determining the content of consciousness (or a report thereof). Finally, it seems to us that in order to properly accommodate consciousness and allow further empirical investigations (e.g. into its relation with other leading models of consciousness in neuroscience, such as the global neuronal workspace, e.g. Baars 1997; Dehaene and Naccache 2001), PEM will have to adopt a somewhat deflationary approach to consciousness, yielding similar results to the position advocated in this paper (i.e. positions similar to those occupied by Carruthers 2000 or Lau 2008), as failing to do this would mean that the framework will likely fail to deliver on its explanatory ambitions.

Acknowledgements Open Access funding provided by Projekt DEAL. We are extremely grateful to several anonymous reviewers, and also to numerous conference audiences, all of whom have provided helpful feedback on earlier drafts. It has been a long time coming—the original idea for this paper was developed when both authors were MSc students at the University of Edinburgh in 2013, and it was first presented at the 2014 Rudolf Carnap Lectures, where Daniel Dennett encouraged us to continue working on it. Since then we have both completed PhDs and started postdoctoral research, and so we should also thank all of our friends, family, and colleagues who have supported us along the way.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included

in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, 218(3), 611–643.
- Ansorge, U., Kunde, W., & Kiefer, M. (2014). Unconscious vision and executive control: How unconscious processing and conscious action control interact. *Consciousness and Cognition*, 27, 268–287.
- Baars, B. J. (1997). In the theatre of consciousness: Global Workspace Theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, 4(4), 292–309. <https://doi.org/10.1093/acprof:oso/9780195102659.001.1>.
- Bayne, T. (2016). VI—Gist! *Proceedings of the Aristotelian Society*, 116(2), 107–126. <https://doi.org/10.1093/arisoc/aow006>.
- Benson, D. F., & Greenberg, J. P. (1969). Visual Form Agnosia. *Archives of Neurology*, 20(1), 82–89.
- Block, N. J. (1990). Inverted Earth. *Philosophical Perspectives*, 4, 53–79.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18, 227–287.
- Block, N. J. (1996). Mental paint and mental latex. *Philosophical Issues*, 7, 19–49.
- Block, N. J. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, 30(5), 481–548.
- Block, N. (2008). Phenomenal and Access Consciousness. *Proceedings of the Aristotelian Society New Series*, 108(2008), 289–317.
- Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, 76(B), 198–211.
- Brown, H. R., & Friston, K. J. (2013). The functional anatomy of attention: A DCM study. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2013.00784>.
- Carey, D. P., Harvey, M., & Milner, A. D. (1996). Visuomotor sensitivity for shape and orientation in a patient with visual form agnosia. *Neuropsychologia*, 34(5), 329–337.
- Carruthers, P. (2000). *Phenomenal Consciousness: A naturalistic theory*. Cambridge: Cambridge University Press.
- Chalmers, D. J. (1996). *The conscious mind*. Oxford: Oxford University Press.
- Chalmers, D. J. (2004). The representational character of experience. *The Future for Philosophy*. <https://doi.org/10.1093/acprof:oso/9780195311105.003.0011>.
- Chalmers, D. J. (2013). How can we construct a science of consciousness? In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences III* (3rd ed., pp. 1111–1120). Cambridge: The MIT Press.
- Chalmers, D. J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, 25(9–10), 6–61.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>.
- Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. New York: Oxford University Press.
- Cohen, J. (2009). *The Red and The Real: An Essay on Color Ontology*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199556168.001.0001>.
- Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Science*, 15(8), 358–364.
- Colombo, M., & Hartmann, S. (2017). Bayesian Cognitive Science, Unification, and Explanation. *The British Journal for the Philosophy of Science*, 68(2), 451–484.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79(1–2), 1–37.
- Dennett, D. C. (1988). Quining qualia. *Philosophy of Mind: Classical and Contemporary Readings*. <https://doi.org/10.1093/acprof:oso/9780198522379.001.0003>.
- Dennett, D. C. (1991). *Consciousness Explained*. London: The Penguin Press.

- Dennett, D. C. (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press.
- Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353(1373), 1245–1255. <https://doi.org/10.1098/rstb.1998.0280>.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222. <https://doi.org/10.1146/annurev.ne.18.030195.001205>.
- Dewhurst, J., & Dołęga, K. (Forthcoming). Attending to the Illusion of Consciousness. In Kammerrer Francoise (Ed.) Symposium on David Chalmers' 'The Meta-Problem of Consciousness, second part—Special issue of the *Journal of Consciousness Studies*.
- Dijkstra, N., van de Nieuwenhuijzen, M. E., van Gerven, M. A. J. (2016). The spatiotemporal dynamics of binocular rivalry: Evidence for increased top-down flow prior to a perceptual switch. *Neuroscience of Consciousness*, 1, niw003. <https://doi.org/10.1093/nc/niw003>.
- Dołęga, K. (2017). Moderate Predictive Processing: Philosophy and Predictive Processing. In W. Wiese & T. Metzinger (Eds.), *Philosophy and Predictive Processing* (pp. 161–179). Frankfurt am Main: MIND Group. <https://doi.org/10.15502/9783958573116>.
- Dołęga, K. (2018). Commentary: M-Autonomy. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2018.00680>.
- Dołęga, K., & Dewhurst, J. (2015). Curtain call at the Cartesian theatre. *Journal of Consciousness Studies*, 22(9–10), 109–128.
- Dołęga, K., & Dewhurst, J. (2019). Bayesian Frugality and the Representation of Attention. *Journal of Consciousness Studies*, 26(3–4), 38–63.
- Dretske, F. (1993). Conscious Experience. *Mind*, 102, 263–283.
- Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: Bradford Books/MIT Press.
- Dreske, F. (1996). Phenomenal externalism, or if meanings ain't in the head, where are qualia? *Philosophical Issues*, 7, 143–158.
- Dretske, F. (2003). How Do You Know You Are Not a Zombie? In B. Gertler (Ed.), *Privileged Access and First Person Authority*. Aldershot: Ashgate Publishing Limited.
- Faivre, N., Mudrik, L., Schwartz, N., & Koch, C. (2014). Multisensory integration in complete unawareness: Evidence from audiovisual congruency priming. *Psychological Science*, 25, 2006–2016. <https://doi.org/10.1177/0956797614547916> (Epub 2014 Sep 30).
- Feldman, H., & Friston, K. J. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2010.00215>.
- Frankish, K. (2016a). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11–12), 65–72.
- Frankish, K. (2016b). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11–12), 256–259.
- Friston, K. J. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, 4(11). (Online), <http://www.fil.ion.ucl.ac.uk/~karl/Hierarchical%20Models%20in%20the%20Brain.pdf> (14 July 2014).
- Friston, K. J. (2010). The free-energy principle: A unified brain theory? *National Review of Neuroscience*, 11(2), 127–138.
- Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS ONE*, 4(7), e6421 (Online).
- Friston, K. J., & Kiebel, S. J. (2009). Predictive coding under the free-energy principle. *Phil. Trans. R. Soc. B*, 364, 1211–1221. <https://doi.org/10.1098/rstb.2008.0300>.
- Gładziejewski, P. (2019). Mechanistic unity of the predictive mind. *Theoretical Psychology (ISTP)*, Online first. <https://doi.org/10.1177/0959354319866258>.
- Graziano, M. S. A. (2013). *Consciousness and the social brain*. New York: Oxford University Press.
- Graziano, M. S. A. (2016). Consciousness engineered. *Journal of Consciousness Studies*, 23, 98–115.
- Harman, G. (1990). The intrinsic quality of experience. *Philos. Perspect.*, 4, 31–52. <https://doi.org/10.2307/2214186>.
- Hobson, J. A., & Friston, K. J. (2014). Consciousness, dreams, and inference. *Journal of Consciousness Studies*, 21(1), 1–27.
- Hobson, J. A., & Friston, K. J. (2016). A Response to Our Theatre Critics. *Journal of Consciousness Studies*, 23(3–4), 245–254.
- Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, 3(APR). <https://doi.org/10.3389/fpsyg.2012.00096>.

- Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.
- Hohwy, J. (2014). The self-evidencing brain. *Noûs*. <https://doi.org/10.1111/nous.12062>.
- Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, 108(3), 687–701. <https://doi.org/10.1016/j.cognition.2008.05.010>.
- Holt, J. (2013). *Blindsight and the nature of consciousness*. Broadview Press.
- Hooker, C. A. (1981). Towards a General Theory of Reduction. Part III: Cross-Categorical Reduction. *Dialogue*, 20(03), 496–529. <https://doi.org/10.1017/S0012217300023593>.
- Humphrey, N. (1974). Vision in a monkey without striate cortex: A case study. *Perception*, 3, 241–255. <https://doi.org/10.1068/p030241>.
- Humphrey, N. (2011). *Soul Dust: The Magic of Consciousness*. New Jersey: Princeton University Press.
- Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. *Perceptual and Motor Skills*, 37(3), 967–979. <https://doi.org/10.2466/pms.1973.37.3.967>.
- Jackson, F. (2003). Mind and Illusion. In A. O’Hear (Ed.), *Royal Institute of Philosophy Supplement*. Cambridge: Cambridge University Press.
- Kentridge, R. W. (2013). Visual attention: Bringing the unseen past into view. *Current Biology*, 23(2), R69–R71.
- Kind, A. (2007). Restrictions on Representationalism. *Philosophical Studies*, 134(3), 405–427.
- Knotts, J. D., Odegaard, B., Lau, H., & Rosenthal, D. (2018). Subjective inflation: Phenomenology’s get-rich-quick scheme. *Current Opinion in Psychology*, 29, 49–55.
- Kolers, P. A., & von Grünau, M. (1976). Shape and color in apparent motion. *Vision Research*, 16(4), 329–335.
- Lamme, V. (2006). Toward a true neural stance on consciousness. *Trends in Cognitive Science*, 10(11), 494–501.
- Lau, H. (2008). A higher order Bayesian decision theory of consciousness. *Progress in Brain Research*, 168(7), 35–48. [https://doi.org/10.1016/S0079-6123\(07\)68004-2](https://doi.org/10.1016/S0079-6123(07)68004-2).
- Limanowski, J., & Friston, K. (2018). ‘Seeing the Dark’: Grounding Phenomenal Transparency and Opacity in Precision Estimation for Active Inference. *Frontiers in Psychology*, Online first: <https://doi.org/10.3389/fpsyg.2018.00643>.
- Lupyan, G., & Clark, C. (2015). Words and the World: Predictive Coding and the Language-Perception-Cognition Interface. *Current Directions in Psychological Science*, 24(4), 279–284.
- Lycan, W. (2019). Representational Theories of Consciousness. *Stanford Encyclopedia of Philosophy*, 2, 1–24.
- Marr, D. (1982). *Vision: A computational approach*. San Francisco: Freeman & Co.
- Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, 5, 39.
- Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., et al. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience*, 8, 825.
- Metzinger, T. (2003). Phenomenal transparency and cognitive self-reference. *Phenomenology and the Cognitive Sciences*, 2(4), 353–393. <https://doi.org/10.1023/B:PHEN.0000007366.42918.eb>.
- Metzinger, T. (2008). Empirical perspectives from the self-model theory of subjectivity: A brief summary with examples. *Progress in Brain Research*, 168, 215–278. [https://doi.org/10.1016/S0079-6123\(07\)68018-2](https://doi.org/10.1016/S0079-6123(07)68018-2).
- Milkowski, M. (2016). A mechanistic account of computational explanation in cognitive science and computational neuroscience. In V. C. Müller (Ed.), *Computing and Philosophy*. Berlin: Springer.
- Milner, A. D., Perrett, D. I., Johnston, R. S., Benson, P. J., Jordan, T. R., Heeley, D. W., et al. (1991). Perception and action in ‘visual form agnosia’. *Brain*, 114(1), 405–428.
- Moore, G. E. (1903). The refutation of idealism. *Mind*, 12, 433–453.
- Norman, L. J., Akis, K., Heywood, C. A., & Kentridge, R. W. (2014). Color constancy for an unseen surface. *Current Biology*, 24, 2822–2826.
- O’Shea, R. P., Kornmeier, J., & Roeber, U. (2013). Predicting Visual Consciousness Electrophysiologically from Intermittent Binocular Rivalry. *PLoS ONE*, 8(10), e76134. <https://doi.org/10.1371/journal.pone.0076134>.
- Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*. New York: Oxford University Press.
- Piccinini, G. (2003). Data from introspective reports: Upgrading from common sense to science. *Journal of Consciousness Studies*, 10(9–10), 141–156.

- Pickering, M. J., & Clark, A. (2014). Getting Ahead: Forward Models and Their Place in Cognitive Architecture. *Trends in Cognitive Sciences*, 18(9), 451–456.
- Prigatano, G. P., & Schacter, D. L. (1991). *Awareness of deficit after brain injury: Clinical and theoretical issues*. Oxford: Oxford University Press.
- Rao, R. P. N., & 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>.
- Schwitzgebel, E. (2007). Do You Have Constant Tactile Experience of Your Feet in Your Shoes? Or Is Experience Limited to What's in Attention? *Journal of Consciousness Studies*, 14(3), 5–35.
- Schwitzgebel, E. (2012a). Self-Ignorance. In Jee Loo Liu & John Perry (Eds.), *Consciousness and the Self*. Cambridge: Cambridge.
- Schwitzgebel, E. (2012b). Introspection, What? In Declan Smithies & Daniel Stoljar (Eds.), *Introspection and Consciousness*. Oxford: Oxford.
- Sergent, C., Wyart, V., Babo-Rebelo, M., Cohen, L., Naccache, L., & Tallon-Baudry, C. (2013). Cueing Attention after the Stimulus Is Gone Can Retrospectively Trigger Conscious Perception. *Current Biology*, 23(2), 150–155.
- Smith, D.W. (2018). Phenomenology. *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/sum2018/entries/phenomenology/>.
- Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, 112, 92–97.
- Thibault, L., Van den Berg, R., Cavanagh, P., & Sergent, C. (2016). Retrospective attention gates discrete conscious access to past sensory stimuli. *PLoS ONE*, 11(2), e0148504.
- Titchener, E. B. (1901). *Experimental psychology: A manual of laboratory practice*, Vol. 1. *Qualitative experiments: Part 2. Instructor's manual*. New York: MacMillan Co.
- Tye, M. (1995). *Ten Problems of Consciousness*. Cambridge, MA: Bradford Books/MIT Press.
- Tye, M. (2003). *Consciousness and Persons*. Cambridge, MA: Bradford Books/MIT Press.
- Tye, M. (2009). *Consciousness Revisited: Materialism without Phenomenal Concepts*. Cambridge, MA: MIT Press.
- 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>.
- Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications*. Oxford: Oxford University Press.
- Wiese, W. (2017). Action is enabled by systematic misrepresentations. *Erkenntnis*, 82, 1233–1252. <https://doi.org/10.1007/s10670-016-9867-x>.
- Wiese, W., & Metzinger, T. (2017). Vanilla PP for philosophers: A primer on predictive processing. In W. Wiese & T. Metzinger (Eds.), *Philosophy and Predictive Processing* (pp. 1–18). Frankfurt am Main: MIND Group.
- Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, 11, R729–R732. [https://doi.org/10.1016/S0960-9822\(01\)00432-8](https://doi.org/10.1016/S0960-9822(01)00432-8).
- Wu, W. (2011). Confronting many-many problems: Attention and agentive control. *Noûs*, 45, 50–76.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.