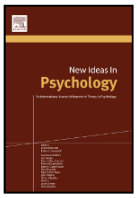
Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: http://ees.elsevier.com



Is predictive processing a theory of perceptual consciousness?

Tomáš Marvan [a,*], Marek Havlík [b]

[a] Institute of Philosophy, Czech Academy of Sciences, Prague, Czech Republic
[b] National Institute of Mental Health, Prague, Czech Republic

ARTICLE INFO

Keywords
Predictive processing
Consciousness
Prerequisites of consciousness
Prediction error minimization
Perceptual hypothesis

ABSTRACT

Predictive Processing theory, hotly debated in neuroscience, psychology and philosophy, promises to explain a number of perceptual and cognitive phenomena in a simple and elegant manner. In some of its versions, the theory is ambitiously advertised as a new theory of conscious perception. The task of this paper is to assess whether this claim is realistic. We will be arguing that the Predictive Processing theory cannot explain the transition from unconscious to conscious perception in its proprietary terms. The explanations offer by PP theorists mostly concern the preconditions of conscious perception, leaving the genuine material substrate of consciousness untouched.

1. Introduction

Predictive processing (PP) is currently one of the most debated theories of brain function. The theory pictures the brain as a hypothesis testing machine matching perceptual hypotheses (priors or prior beliefs) generated by an internal hierarchical model with inputs coming through sensory channels. Hypotheses of the internal model are based on learning as well as "hard-wired" evolutionary constraints (Otten et al., 2017). The mismatch between a hypothesis and the sensory input amounts to "prediction error". Such a mismatch is propagated higher up the hierarchy of the model, until higher-level hypotheses are adjusted accordingly. This process of prediction error minimization (PEM) is concurrently running in the brain on multiple time scales, at various stages of the perceptual hierarchy and in various brain regions where the parts of the internal model are embedded. Organisms capable of acting are not bound to constant passive updating of their internal models. They can act on the world, thus actively changing sensory inputs to match them with aspects of the internal model ("active inference"; see Parr et al., 2019).

A lot of hope is currently put into the PP theory. For instance, under the pressure of current fashion, deep brain networks are being redeveloped along the lines of the PP models (Dora et al., 2018; Lotter et al., 2017). There is also work in progress on the PP analysis of meta-awareness and higher order cognition (Fleming, 2019; Fleming & Daw, 2017). Symptoms that accompany psychosis, such as delusions and hallucinations, are now being reconsidered in light of current PP theories (Adams et al., 2014; Corlett et al., 2019; Sterzer et al., 2018). Influence of the approach can also be documented by recent attempts to re-interpret the significance of neuronal activations captured by fMRI scans (Alink et al., 2010) and the function of EEG oscillations (Heilbron & Chait, 2018) under the prism of the PP theory. Philosophers and neuroscientists increasingly assume that PP will explain perception (Hohwy et al., 2008), attention (Feldman & Friston, 2010) and action (Clark, 2013) in a systematic and unified manner.

In short, in some quarters, PP is expected to become the global theory of brain function (Friston, 2010). This zeal should be somewhat tempered by the fact that many contemporary neurobiological models of perception and cognition do not work with predictive architectures. Not just that: some theorists are openly sceptical, claiming either that there is no evidence for prediction error architecture in the obtained data (Kogo & Trengove, 2015; Philips et al., 2018, p. 8) or that the brain cannot perform the operations hypothesized by the PP theorists (Purves et al., 2015). Other authors embrace some of the PP ideas but accord neural predictions only a limited role in their accounts of brain function (Bullier, 2006; Heeger, 2017). Despite these unsettled questions, the broadness of its explanatory scope, combined with the relative simplicity of its explanatory principles, make the PP theory attractive for many theorists and disciplines.

Must the PP theory explain perceptual consciousness as well?[1] Conscious states represent a large part of an agent's mental life and unfold differently from the non-conscious ones. If PP indeed aspires to be an all-encompassing theory of the brain, to explain "perception and action and everything mental in between" (Hohwy, 2013, p. 1), explanation of consciousness is to be expected from PP proponents. Of course, the PP theory might be taken less ambitiously. For example, it might be taken just as a theory of information flow in cortical circuits which conforms to some first principle assumptions about living systems, such as the free energy principle. Such a theory need not aspire to explain the

* Corresponding author.
E-mail address: marvan@flu.cas.cz (T. Marvan)

https://doi.org/10.1016/j.newideapsych.2020.100837
Received 26 June 2020; Received in revised form 10 November 2020; Accepted 12 November 2020
Available online xxx
0732-118/

difference between conscious and unconscious processing, while still being theoretically exciting. As a matter of fact, though, some authors working in the PP framework advertise PP as a theory that contributes to the general explanation of consciousness. Sometimes it is a matter of indirect but telling hints. Recall Hohwy's remark that PP is expected to explain perception and action and everything mental in between. Presumably, when speaking of perception he intends to cover the cases of *conscious* perception as well, not exclusively the instances of unconscious perception. This is confirmed by Hohwy's other published statements, such as his remark that the PP account of binocular rivalry he developed with Roepstorff and Friston "furnishes the first PP approach to visual *consciousness*" (Hohwy, 2020, p. 7; emphasis in the original), or that we can "discover aspects of consciousness in a general account of brain function" offered by PP (Hohwy, 2015, p. 321). Other authors do not invoke PP's aspiration to become the general account of brain function, but simply proceed to state that PP delivers a new theory of conscious perception. Hobson and Friston (2014, p. 22) boldly assert that consciousness "is not a hard thing to understand, describe, or make hypotheses about — if one associates it with inference based on deeply structured hierarchical (probabilistic) beliefs about sensations"; Wiese and Metzinger (2017, p. 2) claim that if the PP theory is on the right track, "it may provide the means to build new conceptual bridges between theoretical and empirical work on cognition and consciousness"; Clark, Friston and Wilkinson that their version of the PP story attempts to "lay the groundwork for a substantive, but revisionary, account of consciousness itself" (Clark et al., 2019, p. 20); and Drayson (2017, p. 8) that on the PP approach, "conscious perceptual experience is the product of the entire prediction minimization process: it is determined by the interactions between top-down and bottom-up information flow within the entire hierarchy". Perhaps most significantly, Doerig et al. (2020) include PP theory in their list of most influential currently debated theories of consciousness. Assessing the extent to which consciousness is systematically related to the explanatory moves of a PP theory, the central task of this paper, thus seems to be a worthwile exercise.

In the following sections, we focus on contemporary attempts to fit consciousness into the PP explanatory scheme, and critically assess the results reached so far.

2. PP and the difference between unconscious and conscious perception

Does PP has the resources to elucidate, in its own terms, the transition from unconscious to conscious processing of perceptual contents? If the answer is yes, than PP is a theory of consciousness on a par with such worked out approaches as Higher Order Theories (HOT; Brown et al., 2019; Lau & Rosenthal, 2011; Rosenthal, 2005), Global Neuronal Workspace Theory (GNWT; Dehaene, 2014), or attentional (AIR; Prinz, 2012) theories of consciousness. All these theories attempt to explain how mental contents become conscious. The AIR theory gives the following answer: mental contents become conscious when formatted for entry into the working memory. The GNWT gives a different answer: mental contents become conscious by entering into the global neuronal workspace. And according to the HOT theory, mental contents become conscious when appropriately represented by higher-order mental states.

What is the PP theory's answer to this central question? To begin with, some teams report that there is an intimate link between some of the principles postulated by PP theories and the actual contents of conscious perception. For instance, it has been argued that when sensory input is ambiguous, anticipations bias the experienced contents by reducing or ignoring perceptual noise: stimuli are seen as moving, as belonging to a particular object category, etc. in consonance with what is expected to be perceived (de Lange et al., 2018; O'Callaghan et al., 2017; Panichello et al., 2013). In such cases, it might be argued, predictive hypotheses impose a perceptual interpretation on signals that do not contradict it. Furthermore, perceptual expectations improve perceptual metacognition. Sherman et al. (2015) manipulated perceptual expectations by changing the probability that a Gabor stimulus would be presented. Metacognitive sensitivity to the presented stimuli, determined in a standard way as the trial-by-trial correspondence between subjective confidence and objective performance accuracy, increased for expectation-congruent as compared to expectation-incongruent perceptual judgements. Other results indicate different but equally intimate cooperation of PP mechanisms and the mechanisms of consciousness: anticipated stimuli are detected or identified faster than neutral or unpredicted ones (Melloni et al., 2011; Pinto et al., 2015). Although the speed of processing is not necessarily a component of a PP explanation, this might, again, be taken in stride by the PP theory, in the following manner. Stimuli anticipated via internal perceptual models are processed faster and get into conscious stream sooner than neutral and unpredictable ones (although see Mudrik et al., 2011, for results suggesting the opposite conclusion: unexpected stimuli break into consciousness faster than the predictable ones).

However, disambiguation and accelerated processing of stimuli can be in place before conscious processing begins. Using a distinction that got entrenched in consciousness studies (Aru et al., 2012), it is plausible that the mechanisms of disambiguation and acceleration belong to the *prerequisites* of consciousness, not to the genuine neural substrate of conscious processing. Prerequisites of consciousness of the mentioned sort are the mechanisms that gate the perceptual contents or participate in their construction, but then pass them on to other structures that make them conscious.

2.1. PP as a genuine mechanism of consciousness? The constitutive claims

Some theorists offer an account of PP that promises to go beyond prerequisites of consciousness. They propose what might be called a constitutive claim of the relation between predictive processing and conscious perception: predictive processing constitutes the mechanism of transition from unconscious to conscious perception. In this section, we scrutinize their claims in more detail.

A constitutive PP claim tries to fill the gap between unconscious and conscious processing with the conceptual resources proprietary to the PP theory. The tenets of PP relevant in this context are: the nervous system has at its disposal an internal model of anticipatory perceptual hypotheses about the causes of incoming sensory inputs. Sensory signal is dealt with through inference generating perceptual hypotheses with the highest probability, for these hypotheses best explain the incoming signals. If the actually obtaining sensory inputs do not match with anticipations, the model-dependent hypothesis is updated by adapting itself to accommodate prediction errors.

We consider two PP claims that try to illuminate the transition from unconscious to conscious processing. Each of these claims emphasizes a different aspect of predictive processing. The first claim considers consciousness as the result of inferential updating of perceptual hypotheses about the world. The second claim considers consciousness as a mosaic of perceptual hypotheses with the highest overall probability.

2.1.1. Consciousness as inferential updating

Some PP theorists hint that the procedure of making perceptual contents conscious is the *updating* of the perceptual hypotheses via pre-

[1] In the following, we use the term "consciousness" in the sense of perceptual consciousness. We are interested in the mechanisms of the transition between nonconscious and conscious perception of visual, auditory and other contents. We do not enter into debates about "access consciousness" in the sense of Block (1995). We are not interested in how conscious contents become reportable, available to memory and other specialized cognitive systems, etc.

dictive errors propagated through the processing hierarchy. The claim is that what is not updated does not become consciously perceived. In this vein, Hobson and Friston claim consciousness to be "simply the process of optimizing beliefs through inference" (Hobson & Friston, 2016, p. 251), where "beliefs" are to be understood as perceptual hypotheses and "inference" as the "formation of probabilistic perceptual hypotheses by optimizing the sufficient statistics of probability distributions" (Hobson & Friston, 2014, pp. 7). Howhy concurs: "[I]f there is no prediction error to explain away, then there is nothing to be aware of" (Hohwy, 2012, p. 6).

This proposal does not lack initial plausibility. It makes sense to view conscious contents as the results of updates based on the most important and salient events outside of us – events that could not be predicted from the currently applied internal model. Keeping up with unpredicted aspects of reality is crucial, and where better to deal with such unanticipated environmental challenges than directly in conscious experience. However, the view has to face a number of objections. The first and simplest objection is that it does not seem to be the case that every conscious content is a result of evidential update. When I am looking at a familiar book in front of me and do not move myself or the book around, I am not constantly updating my priors on the basis of new evidence; the same perceptual model is applied throughout yet I do not stop consciously seeing the book.[2]

Secondly, some perceptual priors are *stubborn*: they are non-updatable, recalcitrant in the face of new evidence (Yon et al., 2018). Perhaps the most striking example of stubborn perceptual priors is the visual illusion of the Ames Window (Ames, 1951; Bruno et al., 2006). The Ames Window, as the name suggest, is a toy window that has a build-in perspective and rotates slowly in one direction. Windows, prior experiences tell us, are rectangular; the Ames Window is experienced as such even though it is actually a trapezoid. Moreover, it is rotating in one direction at constant speed but it appears to an observer to slow down, stop and reverse the direction of its motion in alternating phases. No matter how much the observers try, they cannot experience the window shape and motion differently, even if a stick or a pen is added as a reference point to help the brain break through the illusion. The stick will unexpectedly pass through the solid frames of the window while the experience of the strangely moving window will stay the same.[3]

The Ames Window represents a set of stubborn, non-updatable perceptual priors concerning the shape and the direction of motion of visual objects. The well-known Hollow Mask Illusion (Gregory, 1973; Grosjean et al., 2012) is another example of the same phenomenon. The observers see a still or a rotating convex face even if they are informed that they are actually looking at the concave side of the mask. Faces are convex and this prior drives the perception. In all such cases, the hardwired or learned constraints on perception are not susceptible to the standard form of evidential updating. On the updating version of the constitutive claim, though, perceptual hypotheses expressing stubborn priors could not become conscious. Such claim appears groundless.

Thirdly, the consciousness-*as*-updating claim is challenged by inferential updates occurring without the reach of consciousness. Examples can be drawn from research on auditory perception. Studies on unexpected stimulus omissions indicate that neural responses in the auditory cortex are shaped by expectations. The neural responses in beat perception, to give one example, rely on pattern expectations. Omission in a predictable rhythm generates a MEG-detectable transient oscillation in the gamma band (Fujioka et al., 2009). Given that gamma activations are tentatively associated with prediction errors (see Heilbron & Chait, 2018, for a critical review), we can treat these early neural responses as prediction error minimization processes.

This sort of predictive updating, though, may not result in conscious perception of the omission. This is clearly seen in congenital amusia subjects. Beat deafness is one variant of this perceptual disorder. Beat-deaf individuals are unable to detect beat omissions and irregularities, or treat irregular beat sequences as regular. However, their early ERP-components such as the mismatch negativity response (MMNr) are typically intact (Mathias et al., 2016). Unconsciously, they remain neurally quite sensitive to unexpected changes in pitch (Moreau et al., 2013, report that amusics neurally respond to violations from expected sound pitch as small as one eighth of a tone; see also Tillmann et al., 2015, pp. 600–601). The detection failures and poor behavioral performance in beat-deaf individuals and other amusics suggest the lack of conscious awareness of the various auditory violations, but their neural updating mechanisms remain intact and active. If we accept that the MMNr is a neural index of updating, we must conclude that updating of auditory hypotheses and discharging of prediction errors may, in congenital amusics, occur entirely unconsciously.

Hobson et al. (2014) mitigate this type of objection by stressing that consciousness pertains only to the higher levels of the presumed hierarchy of generative internal models. For instance, simple motor reflexes successfully discharge prediction errors and revert the body to pre-set proprioceptive equilibrium without the need to consult the consciousness of a subject. This solution acknowledges that consciousness cannot give a place to every sensory hypothesis. If what we consciously experience wasn't heavily filtered, with only a fragment of perceptual hypotheses becoming conscious, we would soon be overloaded with too many updated hypotheses ranging from lowest levels of predictive hierarchy to the highest. Furthermore, as Spratling (2016) notes, many low-level prediction errors will be simply discarded as uninformative. Giving a place to them in consciousness would not only be unnecessary but could lead us astray.

The problem with this hierarchy-based solution is that it is not clear how to demarcate the distinct levels of the supposed hierarchy of generative models, and why should one think that only the higher levels matter for consciousness. The PP hierarchy, it seems to us, does not neatly map onto the more traditional hierarchies such as the hierarchy of processing areas within the visual system, ranging from early regions such as V1 and V2 to the inferior temporal areas and beyond. The PP hierarchy lacks consciousness at its lower levels, but it is simply not true that the contents expressed at the early stages of the standard visual hierarchy are never conscious. We are conscious of low-level features such as textures or colours of objects. Informative specification of the nature of the PP hierarchy of generative models would thus be needed to ascertain that the proposed hierarchical solution works systematically, and is not just an unmotivated ad hoc move.

Finally, the consciousness-*as*-updating story is inconsistent with instances of conscious perception arising without any inferential updates. One such class of instances is constituted by illusory conscious experiences driven by predictive context. An example of the such top-down generated illusion is the well-known Kanizsa illusion, where a triangular shape is perceived even in regions of the image lacking the bottom-up visual evidence for it. The firing activity of neurons in the primary visual cortex with a receptive field on the illusory contour is increased when we look at the Kanizsa triangle. Such illusory experiences are probably evoked by top-down feedback from neurons in higher-order

[2] One of the reviewers points out that some visual aspects of the perceived book, such as the letters and words at its cover, may need to be continually updated via saccades and eye fixations even when we look at the book passively. However, the remaining features such as shape or colour of the book need not be so updated in passive viewing. Thus, our point remains valid: updating is not necessary for all consciously perceived visual features.

[3] One of the reviewers points out that the illusory effect of the Ames window is achieved only when the window is passively viewed on a video screen, i.e., only when the possibility of actively testing the perceptual hypotheses about the movement and the form of the objects is blocked. This limitation, however, is absent in the stubborn prior that we introduce in the next paragraph in the main text.

visual regions with larger receptive fields to deep infragranular layers of the primary visual cortex (Aitken et al., 2020; Kok et al., 2016). Illusory percepts are generated there without any possibility of update from the bottom-up prediction error signals, for these signals (for non-illusory percepts) travel via pyramidal neurons located in the granular, and especially supragranular layers of the primary cortex (Bastos et al., 2012; Mumford, 1992). Provided that inferential updates require a confrontation of predictive feedback and error signals, the consciousness-*as*-update claim cannot be right. Illusory conscious percepts completely bypass the inferential update machinery.

All in all, the version of constitutive claim drawing on the notion of evidential updating does not seem to be successful.

2.1.2. Contents of consciousness fixed by the hypothesis with the highest posterior probability

It could be objected that we read too much into the claims of Hobson, Friston and Hohwy. Maybe they do not want to say that actually occurring evidential updating of hypotheses is what converts unconscious contents to conscious ones. Perhaps they just want to convey the idea that conscious perception is fixed by hypotheses with the highest overall posterior probability. Thus, Howhy writes: "[A]ssume now that conscious perception is determined by the prediction or hypothesis with the highest overall posterior probability" (Hohwy, 2012, p. 4). In a similar vein, Hobson and Friston claim that conscious perception results from the attempts to find "the best (in a Bayes optimal sense) probabilistic explanation for our sensorium" (Hobson & Friston, 2014, p. 7). Howhy (2012) calls the hypothesis with the highest posterior probability the "winning hypothesis"; we follow suit.

This reading bypasses one of the main objections we raised to the previous constitutive claim. According to the present constitutive claim, the book or any other static and unchanging perceptual object will stay in consciousness even though there are no prediction errors to be dealt with. A winning hypothesis that I see a book in front of me is produced by my generative internal model, and this hypothesis alone fixes what I actually see. No updating of my perceptual "belief" is needed. The internal model simply chooses what it takes to be the best possible "guess" of the environmental objects and events, and in some cases, this might be enough to fix conscious contents. But the point of the second constitutive claim is not that evidential updating is not involved in determining what we perceive. Rather, the point is that it is not the updating that drives the transition from unconscious to conscious seeing, but only the winning hypothesis itself.

The internal model-based guessing can diverge from the veridical interpretation of the perceptual scene. Take the well-known phenomenon of fata morgana. Under conditions of temperature inversion the brain hypothesises that the object on the horizon is a body of water, a mountain, a building or even a flying ship. No matter how certain we are that there is nothing like that over there in the distance, the brain will remain persistent in its selection of perceptual contents. Other examples of the same principle are given below in Fig. 1. Most people who have never seen the pictures before will have the following visual experiences: man riding an elephant, a small lying cow, and a herd of very small dinosaurs running from left to right. Time, patience and repeated inspection is needed to realize that the first picture is actually of a well-camouflaged buffalo, the second of a dog with his mazzle up, and the third picture captures a small herd of *Nasua nasua* running in the opposite direction.

Although the majority of the PP literature addresses visual perception, top-down guidance by a preferred perceptual hypothesis is, of course, not limited to it. Let us take an example from the auditory domain, drawn from the studies of so-called phantom words (Deutsch, 2019, ch. 7). Two loudspeakers are put to the left and right of the listener, each loudspeaker repeatedly producing the same sequence of two words or a single two-syllable word, over and over. The sequence

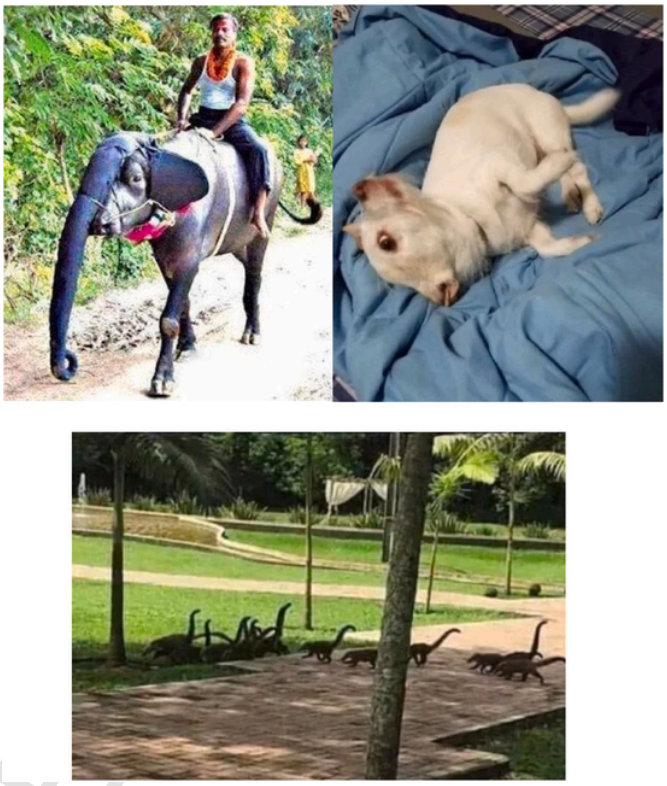


Fig. 1. The brain picks its best possible perceptual guess in accordance with incoming sensory signal and presents such hypothesis as the content of consciousness. – *The pictures are in the common domain and are used under the fair use principle.*

is offset in time in the loudspeakers: when the first word (or the first syllable) is coming from the left loudspeaker, the second word (or the second syllable) is coming from the right loudspeaker (and vice versa). At first, the listener is confronted with a jumble of chaotic sounds. After some time, though, words and phrases are heard. When presented with the word nowhere, people tend to hear phantom words such as window, welcome, love me, run away, no brain, and others (Deutsch, 2019, p. 107). In a PP perspective, this may be interpreted as a process in which the brain draws on the predictive generative model and uses the best guesses to explain away the jumble signal (in the same fashion as in the pictures in Fig. 1). This interpretation can be further supported by the fact that people tend to report hearing phantom words in their native language (Deutsch, 2019, p. 107), and are influenced by the presently meaningful context, such as being on diet and hearing the phantom words "diet coke" or "feel fat" (ibid., p. 108).

Even more dramatic effects of the winning hypothesis can be demonstrated on various psychopathologies such as hallucinations. In these cases, the winning hypothesis is produced independently of the sensory input, or only with a very tenuous reference to this input. As a result, we perceive things that have no counterparts in the external world (see Adams et al., 2014; Sterzer et al., 2018). In a controlled setting, similar effects can to some extent be induced even in healthy subjects (Aru et al., 2018).

Howhy (2012) further specifies that *precision* and *accuracy* are to be taken into account. Precision concerns the varying reliability and noise of the incoming sensory signal. Prediction errors generated by a highly reliable and crisp signal are accorded high precision by the model, while unreliable signals are treated with more caution. Incoming sensory signal may be considered so unreliable that the predictive model is not updated by it; instead, the model sticks to a strong prior. Accuracy is the measure of how well a perceptual hypothesis or a predictive model captures the behaviour of external objects and their

causal relations. In PP terms, more accurate hypotheses and models better supress prediction errors. Howhy adds that although a very complex model might have a broader predictive power than a simpler one, adequately capturing more aspects or details of what is perceived, it is undesirable to have unwieldy and overly complex models. High accuracy therefore needs to be balanced by considerations of simplicity and economic management of resources.

There are no a priori constraints on how much a winning hypothesis needs to be precise or accurate to be consciously perceived. Rather, it is a matter of opportunistic trade-offs between these two dimensions. We typically perceive the contents having both high accuracy and precision, but it is not strictly necessary. We do not loose the ability of conscious perception in contexts involving highly unreliable inputs, and likewise the contents of consciousness are sometimes representationally inaccurate (recall the three pictures at Fig. 1). But the actual manner of how these trade-offs are reached is immaterial to our present purposes. What matters is just the fact that we consciously perceive a hypothesis with what is internally considered the best probabilistic fit, all things considered.

Putting the threads of the previous paragraphs together, we can say that the second constitutive claim promotes the idea that a posterior hypothesis that is taken to embody the currently optimal trade-off between precision and accuracy fixes what we consciously perceive. We will call this idea WHC (standing for "Winning Hypothesis is Conscious"). As in the case of the first constitutive claim, this picture has to face serious objections. The objections we review below use various sources of evidence but all point to the possibility that the winning hypothesis need not reach consciousness.

First, Vetter et al. (2014) experimentally demonstrated that successful predictions in some cases do not terminate in conscious perception. Vetter et al. used the paradigm of apparent motion. In their experiment, the prediction of apparent motion was on some trials generated even if the illusory motion was not perceived consciously. During the experiment, illusory apparent motion was induced by flashing two white squares in rapid succession at different locations. Parallel to this, a target was flashed on the apparent motion trace. On some trials, the flashing of the target was fully synchronized with the illusory motion token, fitting the spatiotemporal prediction. On other trials, targets were flashed out of sync with the token. The subjects were asked to indicate two things: whether they consciously perceived the apparent motion and whether they detected the target. The results show that the detection of targets was better on trials with in time prediction-fitting targets, and that for some (higher) frequencies of the apparent motion-inducing squares this was true for both cases when the subjects consciously perceived the apparent motion *and when they did not*. Vetter et al. conclude that predictions accurately anticipating the unfolding of events are sometimes successfully formed without entering consciousness. Such a possibility, even if it may concern only a minority of perceptual situations, casts doubt on the WHC idea. If some winning hypotheses will not make it to consciousness, just being the winning hypothesis might not be sufficient being perceived consciously (though it might be necessary).

Further evidence against the WHC idea is afforded by neuropsychological deficits such blindsight (Pöppel et al., 1973; Sanders et al., 1974; Weiskrantz et al., 1974). Blindsight subjects with lesions in the V1 visual area are able to partially process stimuli in their blind fields without conscious experience, and to adjust their behaviour accordingly. At least to a certain degree: they partly retain the ability to distinguish colours, shapes, to grab objects, and even to catch them. If encouraged, they can navigate the room without hitting objects. As Dołęga and Dewhurst (2020) remark, the visual system of blindsight subjects seems to form imprecise but sufficiently distinct perceptual hypotheses about the objects present in the blind field. These hypotheses are able to guide behaviour to some extent, and we thus have every right to treat them as the hypotheses with the highest overall posterior probability. Given that these hypotheses are unconscious, blindsight and other similar neuropsychological deficits (such as the unilateral neglect – see Vallar & Perani, 1986) are counterexamples to the WHC idea. Dołęga and Dewhurst conclude that the WHC idea cannot on its own explain the transition from unconscious to conscious perception.

The same conclusion can be reached even without appeal to perceptual pathologies. Recall the view of Friston and Hobson that events at the presumed lower levels of the perceptual hierarchy, such as motor reflexes, are successfully executed without the involvement of consciousness. If we accept this view, it again spells trouble for the WHC idea. It is plausible to view the reflexes as being guided by sufficiently accurate and precise winning hypotheses. The winning hypotheses at the presumed lower levels are therefore dissociated from consciousness. They do not become conscious just in virtue of being winning posteriors. Further conditions need to be met.

Continuous flash suppression (CFS; Tsuchiya & Koch, 2005) is the experimental paradigm similar in some aspects to the more known paradigm of binocular rivalry (Blake & Logothetis, 2002). In both of these inter-ocular paradigms two different stimuli are presented to each eye separately. In case of the CFS, one stimulus (an arrow, for instance) is presented to one eye while the other eye is stimulated with highly salient and constantly changing patterns of shapes ("Mondrians"). Both paradigms are based on the phenomenon of neural rivalry. There is a competition between underlying neural processes in which one of the perceptual versions becomes dominant and enters the stream of consciousness. During binocular rivalry, the versions of both eyes spontaneously alternate in consciousness. During CFS, however, the flashing Mondrians render the pattern presented to the other eye unconscious for long periods. The suppressed pattern is the winning perceptual hypothesis correctly reflecting the input to the respective eye. It can prime responses while being kept entirely out of consciousness (Koivisto & Grassini, 2018; Yang et al., 2014). The fact that it is kept out of consciousness proves, once again, that more is needed for conscious perception than just the emergence of the winning hypothesis.

The arguments marshalled in this section show that the second constitutive claim, based on the winning hypothesis idea, fails too. It cannot explain the transition from unconscious to conscious processing. The processes of which PP speaks can be, at least in some circumstances, completed without terminating in conscious perception. The idea naturally suggests itself: the true mechanisms of consciousness, although closely cooperating with the PP mechanisms, are dissociable from them. PP mechanisms participate in the processes in which the perceptual contents are prepared to be expressed in consciousness, but then pass these contents to other mechanisms that turn them into conscious contents. The latter mechanisms form a set of jointly sufficient neural conditions that need to be met if the contents are to enter the ongoing stream of consciousness, and to remain within it at least for a short period of time. If these mechanisms are not recruited, the contents remain unconscious.

3. PP and other theories of consciousness

The preceding sections show that the PP theory cannot explain, in its proprietary terms, how perceptual contents become conscious. To become a genuine theory of consciousness, the PP theory must be supplemented by new explanatory principles directly relevant for consciousness. Alternatively, it must find a way to closely align itself with a different theory that accounts for consciousness-conferring mechanisms. The latter strategy is, of course, less ambitious than the first one, for the heavy lifting of explaining consciousness is done by this independently formulated theory. Given the absence of the more ambitious proposals of the first kind, though, we will focus on examples of the latter kind of strategy, and offer some critical comments.

It has been suggested that there are interesting points of contact between PP and the Information Integration Theory (Bucci & Grasso, 2017), that PP theory can be augmented with Dennett's theory of Multiple Drafts to account for consciousness (Dołęga & Dewhurst, 2020), or that Prinz's attentional AIR theory and the PP theory both locate conscious contents at the intermediate levels of the perceptual hierarchy (Marchi & Hohwy, 2020). Such attempts to forge links between established theories of consciousness and the PP theory are worthwile, of course, but there is also a danger that PP will align itself with theories that give competing and mutually exclusive accounts of what makes perceptual contents conscious. A better strategy, it seems, is to select one particular theory of consciousness and show how the PP account can join its forces with it. This has been done most extensively with the Global Neuronal Workspace Theory (GNWT), to which we now turn.

GNWT is a leading neurobiological theory of consciousness supported by impressive amount of evidence (for the review of which see Dehaene, 2014). Hohwy (2013) leads the way in sketching how to integrate GNWT and PP into a unified account. GNWT does not in any important way rely on predictions, so Hohwy's proposed extension of it is genuinely novel. It is supposed to work like this. (1) The explanation of how perceptual contents enter the conscious stream is secured by the GNWT itself: contents get conscious by entering the prefronto-parietal neuronal "workspace", and staying within it for at least a short while. By entering the workspace and staying within it, contents become available to various consumer subsystems; this is what makes them conscious. (2) Entry into the workspace is a matter of its non-linear ignition. Dehaene (2009) speculates that ignition is triggered when a threshold of unconscious evidence accumulation for a perceptual state is crossed in the process that involves both bottom-up sensory activation and top-down amplification of the sensory signals from the frontal areas. Hohwy notes that such a proposal might be easily translated into the PP terms.

In particular, Hohwy suggests that ignition of the GNW typically happens in the switch between perceptual and active inference (Hohwy, 2013, p. 214). Active inference is the agent's intervention in the world designed to minimize the prediction error not by adjusting the internal model, but by modifying the sensory input by appropriately acting on the world. The active inference idea modifies the GNWT in that the ignition of a subset of workspace neurons is needed for the winning hypothesis to be made available for various consumer systems specifically in the context of acting. Acting needs to take into account various options, select some course of action among them, and stick to it. When the workspace is ignited, the perceptual hypothesis entering it becomes conscious, ready to guide the behaviour as it unfolds in time. Once in the ignited workspace, the selected hypothesis may drive further descending predictions of the sensory input necessary for adaptive action.[4]

This attempt to tie conscious perception and action so closely together might be criticised in the following manner. Most of the time, our conscious perceptual field contains a vast number of presentations that are irrelevant form the point of view of acting.[5] We consciously see buildings and airplanes in the distance, hear noises around us etc., but do not in any way interact with these buildings, airplanes or noises. The relation to active inference could therefore at best concern only a small subset of global workspace contents, not the totality of them. Launching actions seems neither sufficient nor necessary for perceptual contents to become conscious.

Whyte (2019) attempts to take Hohwy's predictive extension of the GNWT one notch further. Drawing in particular on Hohwy et al. (2008), he asks: What if the global workspace *itself* has a predictive organization? In Hohwy's rendering, the minimization of prediction errors occurs before the contents enter the workspace. Once in the GW, they do not behave according to predictive principles anymore. As we have seen, Hohwy insists that the hypothesis in the GW must remain stable enough to be able to guide specific behaviours. According to Whyte's Predictive Global Neuronal Workspace Theory (PGNWT), the architecture that underwrites the global workspace is continuous with the inferential processes of the preconscious perceptual hierarchy: the global workspace itself is engaged in a process of hierarchical prediction error minimization. In particular, Whyte argues that prediction errors accumulated at the lowest levels of the perceptual hierarchy may continue to influence the global distribution of contents in the workspace (which on the GNWT amounts to their being consciously perceived; see section 4.2. of Whyte's 2019 paper for technical details).

Whyte propounds an interpretation of the literature on auditory oddball paradigms that is consistent with the hypothesis that the global neuronal workspace itself has a PP structure. If further corroborated by future studies, the PGNWT will successfully intertwine perceptual inference and prediction error minimization with the genuine neural substrate of consciousness (provided, that is, that GNWT is the correct theory of consciousness; for recent evidence that this may not be the case, see Silverstein et al., 2015; Scott et al., 2018). Suppose that one day this really happens: the PGNWT is robustly supported by evidence. Still, this would not mean that consciousness can be completely explained by PP principles. The mechanism of content distribution in the global workspace, on which the PNGWT piggybacks, will remain the main explanations of how contents become conscious.

A different option is to expand the PP theory with a metacognitive theory of perceptual consciousness. As Hohwy & Seth, (forthcoming) remark, the PP system can incorporate the metacognitive element in the continuous attempts to refine 'higher-level expected precisions'. That is, the PP system is constantly engaged in high-level monitoring of the fluctuations in precisions of its first-order perceptual signals. The goal is to identify regularities in these fluctuations, so that the system may learn to predict future changes in first-order precisions.

The approach of Hohwy & Seth, is in some respects similar to the ideas of Axel Cleeremans. Cleeremans systematically connects consciousness and learning and gives his theory of conscious awareness a metacognitive twist. The idea is that availability to consciousness depends on the extent to which a first-order sensory representation becomes an object of representation for further systems of representation (i.e., meta-representations). Cleeremans calls this process representational redescription. Through it, unconscious first-order representations become objects of representation by being indexed or targeted by metarepresentations (Cleeremans et al., 2020). The brain learns to meta-represent its own first-order states to itself, to 'make them explicit to itself', and this the basis of consciousness (see also Cleeremans, 2011).

The theory of Cleeremans incorporates metacognitive predictions as one of its elements. The proces of representational redescription essentially involves the ability of the higher-order representations to predict how activations of one part of the brain systematically influence activations in another part of the brain (see the example of Supplementary Motor Area and Primary Motor Cortex activations in Cleeremans, 2011, p. 10). However, this predictive colouring is not what makes this theory a theory of perceptual consciousness. The predictions of how ac-

[4] We note that this is consonant with recent experiments indicating that some contents can influence the generation of new top-down predictions only by first becoming conscious (Meijs et al., 2018).

[5] We speak here about acting in the sense of intentional interventions in the world guided by specific action policies. Eye fixations of external objects and other mostly subliminal behaviors subserving perception are not actions in this sense. They can be used to gather new perceptual information about objects and thus to inform perception via prediction errors. But they have an opposite direction of fit than policies guiding active inference. We use these policies to intervene in the world in order to make it conform to our predictions, not to gather new perceptual information.

tivity in one part of the brain influences activity in another part of the brain (and further downstream) can all be carried non-consciously. It is the emergence of full-blown higher-order representations, not their predictive nature, that accounts for the emergence of conscious experience on this theory. Only when the brain learns to represent its own representational states to itself, it becomes conscious.

4. The phenomenal challenge

Can the PP theory aspire to elucidate why experienced contents have the phenomenal character they do? Hohwy proposes that if we start with conscious experience as we know it intimately from the first person, we can use the PP explanatory framework to account for some of its striking features. First, conscious experience is unified. We normally do not get to consciously perceive disjointed contents. The contents are bound together both at the local level and at the global level. (i) At the global level, all conscious contents are always part of the unified perceptual field. Hohwy's PP theory explains the unified nature of the perceptual field as a direct result of the fact that perceptual inference is geared to action (see the previous section 2.2). Action can only be successful if no more than one of the perceptual hypotheses is selected for uptake into consciousness via active inference (see Hohwy, 2013, chap. 5, for further details). Since we can only act consistently if the selected hypothesis is unified, no other unifying work is needed. (ii) At the local level, colours, shapes, textures etc. of objects are "bound" together; we do not get to perceive colour first and texture later, or colours and textures not attached to the object to which they belong. Again, Hohwy thinks that the bound nature of consciously perceived objectual features springs directly from the way the PP explanation is build. The perceptual hypotheses generated by internal models are bound by their very nature; there is no need for a separate dedicated mechanism that would provide the feature binding. (iii) The third aspect of phenomenal character allegedly amenable to PP treatment is the sophisticated mixture of high-level, relatively stable perceptual features, and lower-level, fast-changing features (constrained by the more stable high-level ones, presumably by some form of a neural feedback). I see a book remaining a book (high-level stable feature) under a lot of perceptual variation (low-level fast-changing features) when I move around while looking at it; its surface colours change, its shape and precise distance from my eyes change etc., but perceptually it remains a book. The PP vision of levels of hierarchical internal generative models seems to fit well with this organization of conscious perception.

On this interesting proposal we have three comments. First, the experiential features (i)–(iii) are *structural* features. They all concern the systematic interrelations or groupings of the various contents we consciously perceive. But phenomenal features are, rather, qualitative: the distinctive subjective "feel" of consciously experienced smells, pains or colours. It is not clear how the predictive processing architecture might help explain such qualitative features and our experience of them. To be fair, a theory of consciousness need not aspire to elucidate the phenomenal aspects of perception. The GNWT is an example of a theory that purports to explain how contents enter the stream of consciousness, without saying anything about how their phenomenality is generated. But we take it that PP is a more ambitious type of theory (see Clark, 2016, p. 239; Hohwy, 2012, p. 9). It promises to illuminate perceptual phenomenology, or at least some of its salient aspects. So far it has not delivered on the promise (although see Dennett, 2015, Clark, 2018, and Clark, 2019, for some initial ideas about how the PP models could tackle some of the qualitative aspects of experience).

Our second comment is that all theories of conscious vision known to us accept that there is a hierarchy of distinct processing levels, ranging from lower to higher ones. Explaining the mixture of high- and low-level perceptual features by appealing to a perceptual hierarchy thus is not an innovation of PP theories (see, e.g., Hochstein & Ahissar, 2002).

Third, Hohwy seems to hold that the structural perceptual features (i)–(iii) only occur at the level of consciousness. But that may not be the case. Starting with (iii), the level-based stratification of perceptual contents: there is ample evidence that we can unconsciously perceive both many low-level phenomena such as colours, brightness, orientation, simple shapes, textures and motion, as well as the higher-level phenomena such as shapes in their semantic aspect, permitting the categorization of objects (Prinz, 2017). Arguments for unconscious feature binding (ii) are equally convincing. Prinz reviews evidence for double dissociation between binding and consciousness (Prinz, 2012, pp. 37f.). Perception might be bound during a completely unconscious perceptual process, such as during episodes of masked priming, while, on the other hand, some instances of conscious perception occur in unbound form. The latter option is documented by cases when the stimuli are presented too quickly to be properly bound together (although they do enter conscious stream), or when the subject is afflicted with a perceptual disorder such as associative agnosia.

It is less certain that the first structural feature of experience, the global unity of the perceptual field, can be present unconsciously. It would be controversial to declare that the whole of the perceptual field can be unified already before its contents reach consciousness. The evidence is very limited so far. Here we only note that Mudrik et al. (2011) present results indicating that subjects are able to integrate perceptual elements into a meaningful scene without the benefit of conscious awareness. Such unconscious unification goes far beyond local binding of perceptual features to objects. Notice also that Hohwy's own explanation of how perceptual hypotheses become conscious (via active inference) seem to presuppose a robust form of global unity at the unconscious level. A perceptual hypothesis can guide action only if unified; no consistent course of action can be derived from a disjointed hypothesis. But if the hypothesis is to trigger the ignition of the global neuronal workspace, and thus become conscious, it must be unified already at the preconscious level. On Hohwy's own account, then, consciousness is not needed for the perceptual field to be unified (at least as much unified as is required by successfully acting on the world).

In sum, if the structural aspects (i)–(iii) of perceptual states do not appear only at the conscious level, but can be in place already before consciousness emerges, we are back with the idea, mooted at the beginning of section 2, that mechanisms realizing such features belong to the category of the prerequisites of perceptual consciousness (see also Aru et al., 2016, p. 8).

5. Whatever next with the PP theory of consciousness?

The arguments of this paper show that optimistic claims of some of the PP theorists about consciousness are unsubstantiated. PP theory cannot explain the transition from unconscious to conscious perception or illuminate the phenomenal dimensions of experience in its proprietary terms, drawing on prediction error minimization or the emergence of the winning posterior hypothesis. It is not clear that PP contributes to the explanation of conscious perception beyond identifying some of the candidate prerequisites of conscious perception. The prerequisites of consciousness are the processes preparing and poising perceptual contents for uptake into awareness, but this uptake is secured by other types of mechanisms. It is these other mechanisms that a genuine theory of consciousness must be attempting to isolate and describe.

As noted, PP theorists are currently exploring ways to join forces with other respectable theories of consciousness, most notably with the GNWT. One project for the immediate future would thus be to strengthen the entwinement of PP and GNWT (or other theory of consciousness). The resulting theory, if empirically convincing, could be so deeply trading in predictive explanations that it would no longer be

possible to object that PP account of consciousness rely too much on non-proprietary ideas. Another project would be to work on a fortified PP theory, immune to the objections summarized in this paper. In either case, a canonical version of PP's take on consciousness would be most welcome. At present, various versions of the PP theory are in circulation, emphasizing different aspects or invoking different explanatory principles (see the differences in accounts of Hobson & Friston, 2014, Hohwy, 2013, and Clark, 2019).[6]

Author statement

Tomáš Marvan: Conceptualization, Writing of the Manuscript.
Marek Havlík: Conceptualization, Writing of the Manuscript.

References

Adams, R.A., Brown, H.R., & Friston, K.J. (2014). Bayesian inference, predictive coding and delusions. AVANT J. Philos.-Interdiscip. Vanguard V, 51–88. doi:10.26913/50302014.0112.0004.

Aitken, F., Menelaou, G., Warrington, O., Koolschijn, R., Corbin, N., Callaghan, M., & Kok, P. (2020). Prior expectations evoke stimulus templates in the deep layers of V1. bioRxiv. doi:10.1101/2020.02.13.947622. 2020.02.13.947622.

Alink, A., Schwiedrzik, C.M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. Journal of Neuroscience, 30, 2960–2966. doi:10.1523/JNEUROSCI.3730-10.2010.

Ames, A. (1951). Visual perception and the rotating trapezoidal window. Psychological Monographs, 65(7). doi:10.1037/h0093600. (whole No. 324), i—32.

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

Aru, J., Rutiku, R., Wibral, M., Singer, W., & Melloni, L. (2016). Early effects of previous experience on conscious perception. Neurosci. Conscious, 9, 1–10. doi:10.1093/nc/niw004. eCollection 2016.

Aru, J., Tulver, K., & Bachmann, T. (2018). It's all in your head: Expectations create illusory perception in a dual-task setup. Consciousness and Cognition, 65, 197–208. doi:10.1016/j.concog.2018.09.001.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., & Friston, K.J. (2012). Canonical microcircuits for predictive coding. Neuron, 76, 695–711. doi:10.1016/j.neuron.2012.10.038.

Blake, R., & Logothetis, N. (2002). Visual competition. Nature Reviews Neuroscience, 3, 13–21.

Block, N. (1995). On a confusion about a function of consciousness. Behavioral and Brain Sciences, 18, 227–287.

Brown, R., Lau, H., & LeDoux, J.E. (2019). Understanding the higher-order approach to consciousness. Trends in Cognitive Sciences, 23, 754–768. doi:10.1016/j.tics.2019.06.009.

Bruno, N., Dell'Anna, A., & Jacomuzzi, A. (2006). Ames's window in proprioception. Perception, 35, 25–30. doi:10.1068/p5303.

Bucci, A., & Grasso, M. (2017). Sleep and dreaming in the predictive processing framework. In Metzinger, T., & Wiese, W. (Eds.), PPP - philosophy and predictive processing. Frankfurt am Main: MIND Group. doi:10.15502/9783958573079¨.

Bullier, J. (2006). What is fed back? In van Hemmen, J.L., & Sejnowski, T.J. (Eds.), 23 problems in systems neuroscience (pp. 130–132). Oxford UP.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36, 181–204. doi:10.1017/S0140525X12000477.

Clark, A. (2016). Surfing uncertainty: Prediction, action, and the embodied mind. Oxford University Press.

Clark, A. (2018). Strange inversions. Prediction and the explanation of conscious experience. In Hubner, B. (Ed.), The philosophy of daniel Dennett. Oxford UP.

Clark, A. (2019). Consciousness as generative entanglement. Journal of Philosophy, 116, 645–662. doi:10.5840/jphil20191161241.

Clark, A., Friston, K., & Wilkinson, S. (2019). Bayesing qualia: Consciousness as inference, not raw datum. Journal of Consciousness Studies, 26, 19–33.

Cleeremans, A. (2011). The radical plasticity thesis: How the brain learns to be conscious. Frontiers in Psychology, 2, 86. doi:10.3389/fpsyg.2011.00086.

Cleeremans, A., Achoui, D., Beauny, A., Keuninckx, L., Martin, J.-R., Muñoz-Moldes, S., … de Heering, A. (2020). Learning to be conscious. Trends in Cognitive Sciences, 24(2), 112–123. https://doi:10.1016/j.tics.2019.11.011.

Corlett, P.R., Horga, G., Fletcher, P.C., Alderson-Day, B., Schmack, K., & Powers, A.R. (2019). Hallucinations and strong priors. Trends in Cognitive Sciences, 23(2), 114–127. doi:10.1016/j.tics.2018.12.001.

Dehaene, S. (2009). Conscious and nonconscious processes: Distinct forms of evidence accumulation. Séminaire Poincaré, XII, 89–114.

Dehaene, S. (2014). Consciousness and the brain: Deciphering how the brain codes our thoughts. Penguin.

Dennett, D. (2015). Why and how does consciousness seem the way it seems? In Metzinger, T., & Windt, J.M. (Eds.), Open MIND: 10(T). Frankfurt am Main: MIND Group. doi:10.15502/9783958570245.

Deutsch, D. (2019). Musical illusions and phantom words: How music and speech unlock mysteries of the brain. Oxford: Oxford University Press.

Doerig, A., Schurger, A., & Herzog, M.H. (2020). Hard criteria for empirical theories of consciousness. Cognitive Neuroscience. doi:10.1080/17588928.2020.1772214.

Dora, S., Pennartz, C., & Bohte, S. (2018). A deep predictive coding network for learning latent representations. bioRxiv. doi:10.1101/278218. 278218.

Dołęga, K., & Dewhurst, J.E. (2020). Fame in the predictive brain: A deflationary approach to explaining consciousness in the prediction error minimization framework. Synthese. doi:10.1007/s11229-020-02548-9.

Drayson, Z. (2017). Modularity and the predictive mind. In Metzinger, T., & Wiese, W. (Eds.), PPP – philosophy and predictive processing. Frankfurt am Main: MIND Group. doi:10.15502/9783958573130.

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. Frontiers in Human Neuroscience, 4. doi:10.3389/fnhum.2010.00215.

Fleming, S.M. (2019). Awareness as inference in a higher-order state space. ArXiv190600728 Q-Bio.

Fleming, S.M., & Daw, N.D. (2017). Self-evaluation of decision-making: A general bayesian framework for metacognitive computation. Psychology Review, 124, 91–114. doi:10.1037/rev0000045.

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

Fujioka, T., Trainor, L.J., Large, E.W., & Ross, B. (2009). Beta and gamma rhythms in human auditory cortex during musical beat processing. Annals of the New York Academy of Sciences, 1169, 89–92. https://doi:10.1111/j.1749-6632.2009.04779.x.

Gregory, R.L. (1973). The confounded eye. In Gregory, R.L., & Gombrich, E.H. (Eds.), Illusion in nature and art (pp. 49–96). London: Duckworth.

Grosjean, M., Rinkenauer, G., & Jainta, S. (2012). Where do the eyes really go in the Hollow-Face Illusion? PLoS One, 7(9), e44706. doi:10.1371/journal.pone.0044706.

Heeger, D.J. (2017). Theory of cortical function. Proceedings of the National Academy of Sciences, 114, 1773–1782. doi:10.1073/pnas.1619788114.

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

Hobson, J., & Friston, K. (2014). Consciousness, dreams, and inference the cartesian theatre revisited. Journal of Consciousness Studies, 21, 6–32.

Hobson, J.A., & Friston, K.J. (2016). A response to our theatre critics. Journal of Consciousness Studies, 23(No. 3–4), 245–254.

Hobson, J.A., Hong, C.C.-H., & Friston, K.J. (2014). Virtual reality and consciousness inference in dreaming. Frontiers in Psychology, 5. doi:10.3389/fpsyg.2014.01133.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. Neuron, 36, 791–804. doi:10.1016/s0896-6273(02)01091-7.

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. Frontiers in Psychology, 3. doi:10.3389/fpsyg.2012.00096.

Hohwy, J. (2013). The predictive mind. Oxford University Press UK.

J. Hohwy Prediction error minimization, mental and developmental disorder, and statistical theories of consciousnesshttps://doi.org/10.7551/mitpress/9780262029346.003.00122015293-324

Hohwy, J. (2020). New directions in predictive processing. Mind & Language. doi:10.1111/mila.12281.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. Cognition, 108, 687–701. doi:10.1016/j.cognition.2008.05.010.

Hohwy, J., Seth, A. Predictive processing as a systematic basis for identifying the neural correlates of consciousness. (in press).

Kogo, N., & Trengove, C. (2015). Is predictive coding theory articulated enough to be testable? Frontiers in Computational Neuroscience, 9. doi:10.3389/fncom.2015.00111.

Koivisto, M., & Grassini, S. (2018). Unconscious response priming during continuous flash suppression. PLoS One, 13. doi:10.1371/journal.pone.0192201.

Kok, P., Bains, L.J., van Mourik, T., Norris, D.G., & de Lange, F.P. (2016). Selective activation of the deep layers of the human primary visual cortex by top-down feedback. Curr. Biol. CB, 26, 371–376. doi:10.1016/j.cub.2015.12.038.

de Lange, F.P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? Trends in Cognitive Sciences, 22(No. 9). doi:10.1016/j.tics.2018.06.002. September 2018.

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. Trends in Cognitive Sciences, 15, 365–373. doi:10.1016/j.tics.2011.05.009.

Lotter, W., Kreiman, G., & Cox, D. (2017). Deep predictive coding networks for video prediction and unsupervised learning. ArXiv160508104 Cs Q-Bio.

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

Mathias, B., Lidji, P., Honong, H., Palmer, C., & Peretz, I. (2016). Electrical brain responses to beat irregularities in two cases of beat deafness. Frontiers in Neuroscience, 10, 40. doi:10.3389/fnins.2016.00040.

Meijs, E.L., Slagter, H.A., de Lange, F.P., & van Gaal, S. (2018). Dynamic interactions between top–down expectations and conscious awareness. The Journal of Neuroscience, February, 28(9), 2318–2327. doi:10.1523/JNEUROSCI.1952-17.2017. 2018 • 38.

Melloni, L., Schwiedrzik, C.M., Muller, N., Rodriguez, E., & Singer, W. (2011). Expectations change the signatures and timing of electrophysiological correlates of

[6] Acknowledgements: This work was created on the basis of cooperation between the Institute … … … …. and the National Institute … … …. . XY, representing the … …..., was supported by the Royal Society of Edinburgh and by the Alexander von Humboldt Foundation. YZ, representing the … … …. ., was supported by CSF (GACR) grant no. 17-23718S, and by the NPU I project no. LO1611 from MEYS CR.

perceptual awareness. Journal of Neuroscience, 31(4), 1386–1396. doi:10.1523/JNEUROSCI.4570-10.2011.
Moreau, P., Jolicoeur, P., & Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: Evidence from event-related potentials. Brain and Cognition, 81(3), 337–344. doi:10.1016/j.bandc.2013.01.004.
Mudrik, L., Breska, A., Lamy, D., & Deouell, L. (2011). Integration without awareness. Psychological Science, 22, 764–770. doi:10.1177/0956797611408736.
Mumford, D. (1992). On the computational architecture of the neocortex. II. Biological Cybernetics, 66, 241–251. doi:10.1007/BF00198477.
Otten, M., Seth, A.K., & Pinto, Y. (2017). A social bayesian brain: How social knowledge can shape visual perception. Brain and Cognition, 112, 69–77.
O'Callaghan, C., Kveraga, K., Shine, J.M., Adams, R.B., & Bar, M. (2017). Predictions penetrate perception: Converging insights from brain, behaviour and disorder. Conscious. Cognition, 47, 63–74. doi:10.1016/j.concog.2016.05.003.
Panichello, M.F., Cheung, O.S., & Bar, M. (2013). Predictive feedback and conscious visual experience. Frontiers in Psychology, 3. doi:10.3389/fpsyg.2012.00620.
Parr, T., Corcoran, A.W., Friston, K.J., & Hohwy, J. (2019). Perceptual awareness and active inference. Neuroscience of Consciousness. doi:10.1093/nc/niz012. Volume 2019, Issue 1, 2019. niz012.
Phillips, W.A., Bachmann, T., & Storm, J.F. (2018). Apical function in neocortical pyramidal cells: A common pathway by which general anesthetics can affect mental state. Frontiers in Neural Circuits, 12, 50. doi:10.3389/fncir.2018.00050.
Pinto, Y., van Gaal, S., de Lange, F.P., Lamme, V.A.F., & Seth, A.K. (2015). Expectations accelerate entry of visual stimuli into awareness. Journal of Vision, 15(8). doi:10.1167/15.8.13. 13, 1–15.
Pöppel, E., Held, R., & Frost, D. (1973). Residual visual function after brain wounds involving the central visual pathways in man. Nature, 243, 295.
Prinz, J. (2012). The conscious brain. USA: OUP.
Prinz, J. (2017). Unconscious perception and the function of consciousness. In Radman, Z. (Ed.), Before consciousness. In search of the fundamentals of mind (pp. 142–163). Imprint Academic.
Purves, D., Morgenstern, Y., & Wojtach, W.T. (2015). Perception and reality: Why a wholly empirical paradigm is needed to understand vision. Frontiers in Systems Neuroscience, 9. doi:10.3389/fnsys.2015.00156.
Rosenthal, D.M. (2005). Consciousness and mind. Oxford University Press.
Sanders, M.D., Warrington, E., Marshall, J., & Wieskrantz, L. (1974). "Blindsight": Vision in a field defect. The Lancet, 303, 707–708.
Scott, R.B., Samaha, J., Chrisley, R., & Dienes, Z. (2018). Prevailing theories of consciousness are challenged by novel cross-modal associations acquired between subliminal stimuli. Cognition, 175, 169–185.
Sherman, M.T., Seth, A.K., Barrett, A.B., & Kanai, R. (2015). Prior expectations facilitate metacognition for perceptual decision. Consciousness and Cognition, 35, 53–56. doi:10.1016/j.concog.2015.04.015.
Silverstein, B.H., Snodgrass, M., Shevrin, H., & Kushwaha, R. (2015). P3b, consciousness, and complex unconscious processing. Cortex, 73, 216–227. December 2015.
Spratling, M. (2016). A review of predictive coding algorithms. Brain and Cognition, 112. doi:10.1016/j.bandc.2015.11.003.
Sterzer, P., Adams, R.A., Fletcher, P., Frith, C., Lawrie, S.M., Muckli, L., … Corlett, P.R. (2018). The predictive coding account of psychosis. Biol. Psychiatry, Mechanisms of Cognitive Impairment in Schizophrenia, 84, 634–643. doi:10.1016/j.biopsych.2018.05.015.
Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital amusia. Handbook of clinical neurology: 129 (pp. 589–605). doi:10.1016/B978-0-444-62630-1.00033-0. (3rd series), ch. 33.
Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. Nature Neuroscience, 8, 1096–1101. https://doi:10.1038/nn1500.
Vallar, G., & Perani, D. (1986). The anatomy of unilateral neglect after right-hemisphere stroke lesions. A clinical/CT-scan correlation study in man. Neuropsychologia, 24, 609–622.
Vetter, P., Sanders, L., & Muckli, L. (2014). Dissociation of prediction from conscious perception. Perception, 43, 1107. doi:10.1068/p7766.
Weiskrantz, L., Warrington, E.K., Sanders, M.D., & Marshall, J. (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. Brain, 97, 709–728.
Whyte, C.J. (2019). Integrating the global neuronal workspace into the framework of predictive processing: Towards a working hypothesis. Consciousness and Cognition, 73, 102763. doi:10.1016/j.concog.2019.102763.
W. Wiese T.K. Metzinger Vanilla PP for philosophers: A primer on predictive processinghttps://doi.org/10.15502/9783958573024 2017
Yang, E., Brascamp, J., Kang, M.-S., & Blake, R. (2014). On the use of continuous flash suppression for the study of visual processing outside of awareness. Frontiers in Psychology, 5. doi:10.3389/fpsyg.2014.00724.
Yon, D., de Lange, F., & Press, C. (2018). The predictive brain as a stubborn scientist. Trends in Cognitive Sciences, 23. doi:10.1016/j.tics.2018.10.003.