Is Predictive Processing a Theory of Consciousness?

Tomáš Marvan[1], Marek Havlík[2]

[1] Institute of Philosophy, Czech Academy of Sciences, Prague

[2] National Institute of Mental Health, Prague

Keywords: Predictive Processing – Consciousness – Predictive Coding – Prerequisites of Consciousness

Abstract Predictive Processing theory, hotly debated today 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 to which extent an explanation of consciousness needs to invoke the principles of the PP theory. We will be arguing that the PP theory mostly concerns the preconditions of conscious perception, leaving the genuine material substrate of consciousness largely untouched. Moreover, insofar as it does speak about consciousness, the PP theory is parasitic on other theories when it comes to illuminating the mechanisms of consciousness.

1. Introduction

Predictive processing (PP) is currently one of the most debated theories of brain function. In this mainly top-down information processing scheme, the brain behaves as a hypothesis testing machine that matches 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 predictive error minimisation (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 (Lotter et al., 2017; Dora et al., 2018). There is also work in progress on the PP analysis of meta-awareness and higher order cognition (Fleming, 2019; Fleming and 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; Sterzer et al., 2018; Corlett et al., 2019). Influence of the approach can also be documented by recent attempts to re-interpret the meaning of neuronal activations captured by fMRI scans (Alink et al., 2010) and the function of EEG oscillations (Heilbron and 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 and 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 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 currently no evidence for prediction error architecture in the obtained data (Kogo and 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 (Heeger, 2017; Bullier, 2006). Despite these unsettled questions and lack of decisive empirical confirmation, 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.

Yet, consciousness, a key topic of philosophy and neuroscience, still lacks its clear place within the PP framework. Given that conscious states represent a large part of an agent's mental life and unfold differently from the non-conscious ones, a PP explanation of consciousness is a must if the PP story indeed aspires to be a completely general account of brain function. Assessing the possible extent to which consciousness is systematically related to the principles postulated by a PP theory is the task of this paper. In the following sections, we approach this topic by focusing on contemporary attempts to fit consciousness into the PP scheme and critically assessing the results reached so far. But first, an important conceptual clarification is needed.

2. Conceptual Clarification: The Senses of "Prediction"

The PP literature often conflates different senses of the term "prediction". In one of these senses, predicting is anticipating or expecting the future flow of either raw sensory stimulation or of processed sensory experience. This temporal sense of predictions is sometimes emphasized by philosophers of cognitive science (see, e.g., Clark, 2013; Clark, 2016, p. 28; Clark, Friston and Wilkinson, 2019, p. 21; Wiese and Metzinger, 2017, p. 3). Predictions as (neural) anticipations

also appear in more technical studies of "predictive coding" within the domain of computational neuroscience (Murray et al., 2002; Rao and Ballard, 1999; Schwiedrzik and Freiwald, 2017; Spratling, 2016). In the neurocomputational setting, the emphasis is mainly on the effectiveness of the neural transmission between adjacent levels in the perceptual hierarchy. This effectiveness is enabled by the mixture of feedback (predictions) and feedforward (predictive errors) processing steps. It is not clear how this hangs together with the more sweeping claims of the philosophers, but at least the emphasis on the anticipatory nature of predictions is shared in both approaches.

A different sense of "prediction" frequently invoked by PP theorists concerns the so-called "inverse problem" of perception (see Pizlo, 2001, and Spratling, 2016). Here the term "prediction" is used in a distinctly atemporal way. The inverse problem is this: the perceptual system needs to reconstruct the distal stimuli given only the proximal "sensory barrage" (Clark, 2015, p. 5). How does the perceptual system do this? The PP theory suggests that the system applies perceptual inferences to incoming sensory inputs to reconstruct their hidden distal causes in the external world. This reconstruction draws on the continually updated generative model and allows us to become perceptually aware of objects in front of us. Such "predictions" are geared not to any future states of our sensory pathways, but to what is happening in the present moment. This view of perception goes back to the theories of vision of von Helmholtz and beyond (Hatfield, 2002). The brain is using previous statistical learning to come up with the most suitable interpretation of the sources of events currently occurring "at the nerves", as Helmholtz (1867, p. 430) put it. But the inferences need not rely on any anticipatory activity; only on the "immutable laws" of perception (Helmholtz, 1855, p. 100). In its contemporary forms, the view is infused with the Bayesian vision of evidential updating (Penny, 2012). Hohwy, a prominent PP theorist, is developing his inferential theory along these lines, drawing on work of Helmholtz and Friston (see the overview in Hohwy, 2013; see also Friston, 2012).

The talk of "predictions" in this latter sense is somewhat misleading. Predictions-as-inferences predominantly concern the long-term, stable features of our perceptual models embedded in perceptual priors. For example, we have long-term expectations about depth and colour in different parts of our visual field (Hohwy, 2013, p. 33) or of light coming from above of the visual scene (Penny, 2012, p. 8). These long-term "expectations" are importantly different from the fast changing expectations about how the sensory flow will change in the next instant. However, it is not entirely clear how to integrate these two kinds of predictions into a single theory of perceptual organization. While the perceptual inferences could perhaps be somehow embedded in a dynamical system that is anticipating its future sensory states, they can equally well be triggered independently of any such anticipations. This is most clearly the case in perceptual situations involving genuinely unexpected events.

To mark the differences betwen the two kinds of predictions, we prefer to speak about inferences when the atemporal sense is clearly at play, and about predictions when the anticipatory sense is involved. It seems that whereas philosophers are more interested in the inverse problem of perception, neuroscientists devote more time to the piecemeal studies of neural anticipatory

feedback amongst the presumed levels of the sensory hierarchy. However, clear-cut disambiguation is often difficult and the sense of "prediction" preferred by a particular author is not really clear.

3. The Present-day PP Theories and Consciousness

PP is nowadays often touted as a brand new theory of consciousness, on a par with such worked out approaches as Higher Order Theories (HOT; Brown et al., 2019; Lau and Rosenthal, 2011; Rosenthal, 2005), Global Neuronal Workspace (GNW; Dehaene, 2014) or attentional (AIR; Prinz, 2012) theories of consciousness. However, the details of such a theory are lacking. If PP is to become a theory of consciousness, it must address the central question of consciousness science: what makes some of our mental contents conscious? To give some examples, AIR theory gives the following answer: mental contents become conscious when formatted for entry into the working memory. GNW theory gives a different answer: mental contents become conscious by entering into the global neuronal workspace. And according to 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? Some PP theorists see consciousness as the final result in the series of perceptual processing steps. Thus, Hohwy (2013) portrays consciousness as the "upshot" or "conclusion" of perceptual inferences. Similarly, Melloni (2015) takes consciousness to be the "outcome" or "result" of such inferences. This might suggest that the neural inferential machinery is directly involved in processes that make mental contents conscious. Such conclusion, though, would be premature. The processes of perceptual inference are widely considered to be unconscious. Already in the Helmholtzian account of visual perception the perceptual inferences happen before the percept enters consciousness: they are "unconscious conclusions" (*unbewusste Schlüsse*; Helmholtz, 1867, p. 430). As noted, the modern versions of this story add that perceptual inferences are performed in a manner that approximates the principles of Bayesian evidential updating in uncertain surroundings, and stress the hierarchical nature of internal "generative models". Nonetheless, the inferences remain unconscious.

3.1 PP as a Prerequisite of Consciousness

Given that perceptual inferences are unconscious, one way of fixing the relationship between PP and consciousness is to say that inferential processes prepare perceptual contents for uptake into consciousness. A distinction that got entrenched in consciousness studies (Aru et al., 2012) might thus be used to formulate the following hypothesis: the principles of perceptual organisation that PP theory describes belong to the *prerequisites* of consciousness, not to the genuine neural

substrate of consciousness. Prerequisites of consciousness are the neural mechanisms that participate on constructing the perceptual contents but then pass them on to other structures that make them conscious. According to some recent studies, the PP neural machinery seems to utilize several such content-preparing mechanisms. For instance, it has been argued that when sensory input is ambiguous, anticipations bias the contents of awareness 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 (Panichello et al., 2013; de Lange et al., 2018; O'Callaghan et al., 2017). 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 unpredictable stimuli (Melloni et al., 2011; Pinto et al., 2015; although see Mudrik et al., 2011, for results suggesting unexpected stimuli break into consciousness faster than the predictable ones), and in some circumstances, whether a stimulus is uptaken into consciousness at all may depend on valid expectations (Meijs et al., 2018). In such cases, PP mechanisms perform a brisk triage of possible contents of consciousness, and then push the selected ones for preferential conscious processing. In various psychopathologies such as hallucinations, though, the involvement of PP seems to be even more radical in going beyond triage and speed delivery into consciousness. In these cases, PP *sculpts* these contents independently of the sensory input, or only with a very tenuous reference to this input. It pushes into consciousness even the contents that do not have their counterparts in the external world (see Adams et al., 2014; Sterzer et al., 2018). In a controlled setting, similar effects can be induced even in healthy subjects (Aru et al., 2018).

But even if such results are generally accepted as valid, selection and disambiguation of stimuli, speeding up of their entry into the stream of consciousness, etc. can in principle be in place before conscious processing begins. In other words, it would seem that PP theory itself cannot explain why some contents are conscious and others aren't. It's mechanisms are on a par with other prerequisites of consciousness such as, for instance, the phase of intrinsic brain activity (Sergent and Naccache, 2012, p. 94; Wyart and Sergent, 2009; Northoff, 2014, ch. 15) which can serve as an early predictor of whether a percept will enter conscious awareness or not. Note also that even if we place predictive processing amongst the prerequisites of consciousness, part of the unconscious sensory processing falling under the PP umbrella will nonetheless have no relevance to preconscious mechanisms and eventually to consciousness itself. This concerns especially the events at the lowest levels of the perceptual hierarchy. It might be claimed that visual predictive processing starts at the retina. Still, not every sensory input that impinges on it can be labelled as a prerequisite of consciousness. Many low-level prediction errors will be discarded as uninformative (Spratling, 2016). As Wiese and Metzinger (2017, p. 3) remark, "the contents of phenomenal experience are only part of what is, according to PP, generated through the hierarchically organized process of prediction error minimization (most contents will be unconscious)." In other words, what we consciously see is heavily filtered. Consciousness cannot give a place to every bit of sensory information reaching the body, or generating within it. It would be overloaded with perceptual hypotheses with the highest posterior probability.

3.2 PP as a Genuine Substrate of Consciousness?

Some contemporary theorists seem to think that relegating PP mechanisms to the mere prerequisites of consciousness is too timid. They propose what might be called a constitutive reading of the relation of PP to consciousness: PP machinery is not just intimately connected with the genuine mechanisms of consciousness, but is actually a part and parcel of these mechanisms.

3.2.1. The Constitutive Reading

It is not difficult to find expressions of the constitutive reading in the PP literature, such as this passage from Lamme:

> "(…) consciousness is the result of the unconscious inferential processes. Previous knowledge and experience (the priors) play an important role, but they are combined with current input to produce the posterior, which is conscious sensation" (Lamme, 2015, p. 2),

or this from Hohwy:

> "With this in mind, assume now that conscious perception is determined by the prediction or hypothesis with the highest overall posterior probability (…) conscious perception is determined by the strongest "attractor" in the free energy landscape (…)." (Hohwy, 2012, p. 4)

But how is this supposed to work? What is the procedure by which the best perceptual hypothesis becomes conscious? The simplest way would be to say that the winning hypothesis becomes conscious just by being the hypothesis with the highest posterior probability. But if this is all that can be said in favour of the constitutive reading, the argument is not very convincing. However, some PP theorists hint that the procedure making perceptual contents conscious is the *updating* of the perceptual hypotheses on the basis of predictive errors propagated through the hierarchy. What is not so updated does not become consciously perceived. Howhy again:

> "When all the surprise is dealt with, prediction and model revision should cease. If it is also impossible to do further selective sampling then conscious perception of the object in question should cease. This follows from the idea that what we are aware of is the 'fantasy' generated by the way current predictions attenuate prediction error; *if there is no prediction error to explain away, then there is nothing to be aware of*." (Hohwy, 2012, p. 6; emphasis added)

The same view is espoused by Hobson and Friston who write:

> "We were working at a rather simple (and formal) level in which *consciousness is simply the process of optimizing beliefs through inference*. Implicit in this argument is equivalence between probabilistic beliefs and the products or phenomena of consciousness." (Hobson and Friston, 2016, p. 251; emphasis added)

This reading stresses the role of (sufficiently precise) prediction errors. But the idea is not that the errors themselves become the contents of consciousness. If the agent's consciousness consisted solely of predictive errors, he would be constantly struggling with a high degree of entropy. Such overwhelming entropy would lead the agent to lock himself in the unchanging Dark Room (Friston et al., 2012) and stop exploring his environment altogether. Fortunately, this is not the case. Rather, the contents of consciousness are the perceptual hypotheses informed, not constituted by predictive errors. And according to the version of constitutive reading we are considering, only the actually updated perceptual models count.

Three straightforward objections against the view that conscious perception needs evidential updating come to mind. The first is that, intuitively, it does not seem to be the case that every conscious content is a result of evidential update. When I am looking at a 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.

The second objection is that some priors are "stubborn": they are non-updatable, recalcitrant in the face of new evidence (Yon et al., 2019). A well-known example is the expectation that light comes form above in the perceptual scene. This is a hardwired constraint on perception, not susceptible to the standard form of evidential updating. On the updating version of the constitutive reading, though, perceptual hypotheses expressing stubborn priors don't become conscious. Such claim appears groundless.

The third objection is that, as already noted, not every successful update of the the internal model driven by predictive error will result in conscious perception. Consciousness is selective and has a limited capacity. In contrats, internal models are presumably explaing away prediction errors across all levels of the perceptual hierarchy. The conclusion is, again, that evidential updating does not appear to be the mechanism that confers consciousness on perceptual contents.

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

3.2.2. PP and the Mechanisms of Entry into Consciousness

The preceding section suggests that the PP theory cannot both explain the construction of perceptual contents and how they become conscious. Many of the perceptual contents that are processed by the PP mechanisms do not reach consciousness; consciousness is not an automatic fallout of their activity. This would be in line with the view that is gaining increasing support among consciousness researchers: the mechanisms conferring consciousness on perceptual contents are not intrinsic to the mechanisms constructing the contents. Rather, much of the available evidence points to the contrary conclusion: these two mechanisms operate in substantively different ways. Although it is challenging to distil the two mechanisms experimentally, a growing number of authors stresses the need for such a distinction (Prinz, 2012; Mehta and Mashour, 2013; Bachmann and Hudetz, 2014; Marvan and Polák, 2017; Phillips et al., 2018; Aru et al., 2019). In this perspective, conscious perception is a result of the interaction between content mechanisms and consciousness mechanisms. The latter form a set of jointly sufficient neural conditions 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.

Unless our preceding diagnosis is misguided, PP theory alone cannot constitute a genuine theory of conscious perception. On its own, it cannot explain how perceptual contents become conscious. Of course, it might still remain on the right track regarding the content mechanisms. The genesis of perceptual contents might involve hierarchical inference, top-down predictive feedback and evidential optimisation. In section 3.1 we noted that PP could be involved in triage and disambiguation of stimuli, speeding up of their entry into consciousness etc. Perhaps, as PP theorists hope, this catalogue could be expanded so as to involve PP in all aspects of content preparation. Still, all such aspects would remain in the category of the prerequisites of conscious perception, of its necessary but not sufficient neural preconditions.

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 offers the account of consciousness-conferring mechanisms. The latter strategy is, of course, far 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 one example of the latter kind of strategy, and offer a critical comment.

In the spirit of the latter approach, Hohwy (2013) tries to integrate PP with the Global Neuronal Workspace Theory (GNWT), a leading neurobiological theory of consciousness supported by impressive amount of evidence (for the review of which see Dehaene, 2014). 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. Hohwy notes that such a proposal might be easily translated into 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 predictive error not by adjusting the internal generative 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. Ignition of the global workspace seems fit for this purpose. When ignited, the perceptual hypothesis becomes conscious, ready to guide the behaviour as it unfolds in time; it serves as the best prediction error minimizer for the time being. Once in the ignited workspace, the selected hypothesis may drive further descending predictions of the sensory input deemed necessary for action.[1]

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 completely irrelevant form the point of view of acting. We consciously see buildings and aeroplanes in the distance, hear noises around us etc., but do not in any way interact with these buildings, aeroplanes or noises. The relation to active inference could therefore at best concern only a small subset of conscious contents, not the totality of them. Launching actions thus seems neither sufficient nor necessary for contents to become conscious.

Whyte (2019) tries a different tack. He attempts to take Hohwy's GNWT extension 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 predictive errors occurs before the contents enter the workspace. According to Whyte's Predictive Global Neuronal Workspace (PGNW), the architecture that underwrites the global workspace is continuous with the preconscious perceptual hierarchy. The global workspace itself is engaged in a process of hierarchical predictive error minimisation.

Whyte reviews the literature consistent with the hypothesis that the global neuronal workspace has a PP structure. If further corroborated by future studies, the PGNW theory will successfully intertwine perceptual inference and predictive error minimization with the genuine neural substrate of consciousness (provided that GNWT is the correct theory of consciousness; for recent evidence

[1] 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). We could say that some predictions are in this sense the *consequences* of conscious perception.

that this may not be the case, see Silverstein et al., 2015; Scott et al., 2018). Suppose that one day this really happens: the GNWT is robustly supported by evidence. Still, that would not mean that consciousness can be completely explained by PP principles. The mechanism of content distribution in the global workspace, on which the PNGW theory piggybacks, will remain the main explananans of how contents become conscious. The promise that the PP theory will become a global theory of brain function is not nearly made good on.

4. The Phenomenal Challenge

What about the so-called phenomenal dimensions of consciousness? Can the PP theory aspire to elucidate why experienced contents have the phenomenal character they do? Hohwy admits that one could in principle implement a prediction error minimizing machine that would lack consciousness altogether (Hohwy, 2012, p. 5, fn. 4). This does not stop him from proposing 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 3.2.2). Action can only be successful if 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 belongs. 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 reality-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 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 (lower-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 internal generative models seems to fit well with this hierarchical organization of conscious perception.

On this interesting proposal we have two 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 Global Workspace theory is an example of a theory that purports to explain how contents enter the stream of consciousness, without saying anything about their phenomenality. 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 phenomenology, but so far it has not delivered on the promise (although see Dennett, 2015, and Clark, 2018, for some initial ideas about how the PP models could tackle at least some of the qualitative aspects of experience).

Our second comment is that 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, and 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 conscious awareness. Such unconscious unification goes far beyond local binding of perceptual features to objects. Note 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 seriously disarrayed hypothesis. But if the hypothesis is to trigger the ignition of the global 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).

To sum up, if the structural aspects (i)–(iii) of perceptual contents do not appear only at the conscious level, but can be in place already before consciousness emerges, we are back with the

idea that mechanisms realizing such features belong to the category of prerequisites of consciousness, not to the genuine neural substrate of consciousness.

5. Whatever Next with the PP Theory of Consciousness?

The community of consciousness researchers needs to pause and take stock of the explanatory power and scope of the PP theory as a theory of conscious perception. So far, the enthusiastic claims of its supporters contrast with the fact that the explanation of how contents become consious need not invoke the key notions of the PP theory such as perceptual inference or predictive error minimization. In fact, the PP theory seems to be focusing mainly (or entirely) on the prerequisites of conscious perception: on the various ways the perceptual contents are prepared and poised for uptake into awareness. But these contents, including the best predictive hypotheses, might not become conscious after all. For that to happen, other type of mechanism seems to be required.

The PP theorists should indicate whether the plan is to further integrate the PP theory with other self-standing theories of consciousness, such as the Global Neuronal Workspace theory, or whether a truly predictive theory of consciousness is forthcoming. Such a theory would attempt to explain, in its own vocabulary, what makes perceptual contents conscious. Another pressing matter is the issue of insufficient evidence for the PP models. This does not concern just the models of consciousness, but more generally the models of perception. The link between the sweeping claims of PP theorists such as Clark and Hohwy and the piecemeal and detailed evidence found in the predictive coding studies of computer neuroscience needs to be clarified. In the same vein, explanations are needed of how to cast the technical terms of PP theories such as “inference” or “perceptual hypothesis” in naturalistic, preferably neuronal terms.

Bibliography

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

Alink, A., Schwiedrzik, C.M., Kohler, A., Singer, W., Muckli, L., 2010. Stimulus Predictability Reduces Responses in Primary Visual Cortex. J. Neurosci. 30, 2960–2966. https://doi.org/10.1523/JNEUROSCI.3730-10.2010

Aru, J., Bachmann, T., Singer, W., Melloni, L., 2012. Distilling the neural correlates of consciousness. Neurosci. Biobehav. Rev. 36, 737–746. https://doi.org/10.1016/j.neubiorev.2011.12.003

Aru, J., Tulver, K., Bachmann, T., 2018. It's all in your head: Expectations create illusory perception in a dual-task setup. Conscious Cogn. 65, 197-208. https://doi.org/10.1016/j.concog.2018.09.001
Aru, J., Mototaka, S., Rutiku, R., Larkum, M.E., Bachmann, T., 2019. Coupling the State and Contents of Consciousness. Front. Syst. Neurosci. 13, 1–9. https://doi.org/10.3389/fnsys.2019.00043
Bachmann, T., Hudetz, A., 2014. It is time to combine the two main traditions in the research on the neural correlates of consciousness: C = L x D. Front. Psychol. 5, 1–13. https://doi.org/10.3389/fpsyg.2014.00940
Brown, R., Lau, H., LeDoux, J.E., 2019. Understanding the Higher-Order Approach to Consciousness. Trends Cogn. Sci. 23, 754–768. https://doi.org/10.1016/j.tics.2019.06.009
Bullier, J., 2006. What Is Fed Back? In J. L. van Hemmen & T J. Sejnowski, 23 Problems in Systems Neuroscience. Oxford UP, 130–132.
Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36, 181–204. https://doi.org/10.1017/S0140525X12000477
Clark, A., 2015. Predicting Peace: The End of the Representation Wars. A Reply to Michael Madary. In T. Metzinger & J. M. Windt (Eds). Open MIND: 7(R). Frankfurt am Main: MIND Group. https://doi:10.15502/9783958570979
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 In B. Hubner (ed.), The Philosophy of Daniel Dennett. Oxford UP.
Clark, A., Friston, K., Wilkinson, S., 2019. Bayesing qualia: Consciousness as inference, not raw datum. J. Conscious. Stud. 26, 19–33.
Corlett, P.R., Horga, G., Fletcher, P.C., Alderson-Day, B., Schmack, K., Powers, A.R., 2019. Hallucinations and Strong Priors. Trends Cogn Sci. Feb;23(2):114-127. https://doi.org/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.
de Lange, F.P., Heilbron, M., Kok, P., 2018. How Do Expectations Shape Perception? Trends in Cognitive Sciences, September 2018, Vol. 22, No. 9. https://doi.org/10.1016/j.tics.2018.06.002
Dennett, D., 2015. Why and How Does Consciousness Seem the Way it Seems? In T. Metzinger & J. M. Windt (Eds). Open MIND: 10(T). Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958570245
Dora, S., Pennartz, C., Bohte, S., 2018. A Deep Predictive Coding Network for Learning Latent Representations. bioRxiv 278218. https://doi.org/10.1101/278218
Feldman, H., Friston, K., 2010. Attention, Uncertainty, and Free-Energy. Front. Hum. Neurosci. 4. https://doi.org/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. Psychol. Rev. 124, 91–114. https://doi.org/10.1037/rev0000045
Friston, K., 2012. The history of the future of the Bayesian brain. NeuroImage 62, 1230–1233. https://doi.org/10.1016/j.neuroimage.2011.10.004
Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11, 127–138. https://doi.org/10.1038/nrn2787
Friston, K., Thornton, C., Clark, A., 2012. Free-Energy Minimization and the Dark-Room Problem. Front. Psychol. 3. https://doi.org/10.3389/fpsyg.2012.00130
Hatfield, G., 2002. Perception as Unconscious Inference. In Heyer, D., Mausfeld, R. (Eds.), Perception and the Physical World: Psychological and Philosophical Issues in Perception. John Wiley and Sons, pp. 113–143.
Heeger, D.J., 2017. Theory of cortical function. Proc. Natl. Acad. Sci. 114, 1773–1782. https://doi.org/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. https://doi.org/10.1016/j.neuroscience.2017.07.061
Helmholtz, H. von. 1855. Über das Sehen des Menschen. In Helmholtz, Vorträge und Reden. 4th edition, Vol .1. 1896. Braunschweig: Vieweg.
Helmholtz, H. von, 1867. Handbuch der physiologischen Optik. Leipzig : Leopold Voss.
Hobson, J.A., Friston, K.J., 2016. A Response to Our Theatre Critics. Journal of Consciousness Studies, 23, No. 3–4, 245–54.
Hohwy, J., 2012. Attention and Conscious Perception in the Hypothesis Testing Brain. Front. Psychol. 3. https://doi.org/10.3389/fpsyg.2012.00096
Hohwy, J., 2013. The Predictive Mind. Oxford University Press UK.
Hohwy, J., Roepstorff, A., Friston, K., 2008. Predictive coding explains binocular rivalry: An epistemological review. Cognition 108, 687–701. https://doi.org/10.1016/j.cognition.2008.05.010
Kogo, N., Trengove, C., 2015. Is predictive coding theory articulated enough to be testable? Front. Comput. Neurosci. 9. https://doi.org/10.3389/fncom.2015.00111
Lamme, V.A.F., 2015. Predictive Coding Is Unconscious, so that Consciousness Happens Now. In T. Metzinger & J. M. Windt (Eds.), Open MIND: 22(C). Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958571105
Lau, H., Rosenthal, D., 2011. Empirical support for higher-order theories of conscious awareness. Trends Cogn. Sci. 15, 365–373. https://doi.org/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.
Marvan, T., Polák, M., 2017. Unitary and dual models of phenomenal consciousness. Conscious. Cogn. 56, 1–12. https://doi.org/10.1016/j.concog.2017.09.006

Mehta, N., Mashour, G.A., 2013. General and specific consciousness: a first-order representationalist approach. Front. Psychol. 4, 407. https://doi.org/10.3389/fpsyg.2013.00407
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, 2018 • 38(9):2318–2327. https://doi.org/10.1523/JNEUROSCI.1952-17.2017
Melloni, L., Schwiedrzik, C. M., Muller, N., Rodriguez, E.
& Singer, W. (2011). Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. The Journal of Neuroscience, 31(4), 1386–1396. https://doi.org/10.1523/JNEUROSCI.4570-10.2011
Melloni, L. (2015). Consciousness as Inference in Time. A Commentary on Victor Lamme. In T. Metzinger & J. M. Windt (Eds.), Open MIND: 22(C). Frankfurt am Main: MIND Group. https://doi.or/10.15502/9783958570566
Mudrik, L., Breska, A., Lamy, D., Deouell, L., 2011. Integration Withut Awareness. Psychol. Sci. 22, 764–70. https://doi.org/10.1177/0956797611408736
Murray, S.O., Kersten, D., Olshausen, B.A., Schrater, P., Woods, D.L., 2002. Shape perception reduces activity in human primary visual cortex. Proc. Natl. Acad. Sci. 99, 15164–15169. https://doi.org/10.1073/pnas.192579399
Northoff, G., 2014. Unlocking the Brain. Volume 2: Consciousness. Oxford UP.
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. Cogn. 47, 63–74. https://doi.org/10.1016/j.concog.2016.05.003
Otten, M., Seth, A. K., and Pinto, Y. 2017. A social Bayesian brain: how social knowledge can shape visual perception. Brain and Cognition, 112. pp. 69–77.
Panichello, M.F., Cheung, O.S., Bar, M., 2013. Predictive Feedback and Conscious Visual Experience. Front. Psychol. 3. https://doi.org/10.3389/fpsyg.2012.00620
Parr, T., Andrew W Corcoran, Karl J Friston, Jakob Hohwy, 2019. Perceptual awareness and active inference. Neuroscience of Consciousness, Volume 2019, Issue 1, 2019. niz012, https://doi.org/10.1093/nc/niz012
Penny, W., 2012. Bayesian Models of Brain and Behaviour [WWW Document]. ISRN Biomath. https://doi.org/10.5402/2012/785791
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. Front. Neural Circuits, 02 July 2018. https://doi.org/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):13, 1–15. https://doi.org/10.1167/15.8.13
Pizlo, Z., 2001. Perception Viewed as an Inverse Problem. Vision Research, 41(24), 3145–3161. https://doi.org/10.1016/S0042-698900173-0
Prinz, J., 2012. The Conscious Brain. OUP USA.

Prinz, J., 2017. Unconscious Perception and the Function of Consciousness. In Radman, Z. (ed.), Before Consciousness. In Search of the Fundamentals of Mind. Imprint Academic, 142–163.
Purves, D., Morgenstern, Y., Wojtach, W.T., 2015. Perception and Reality: Why a Wholly Empirical Paradigm is Needed to Understand Vision. Front. Syst. Neurosci. 9. https://doi.org/10.3389/fnsys.2015.00156
Rao, R., Ballard, D., 1999. Predictive Coding in the Visual Cortex: a Functional Interpretation of Some Extra-classical Receptive-field Effects. Nat. Neurosci. 2, 79–87. https://doi.org/10.1038/4580
Rosenthal, D.M., 2005. Consciousness and mind. Oxford University Press.
Schwiedrzik, C.M., Freiwald, W.A., 2017. High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing Hierarchy. Neuron 96, 89-97.e4. https://doi.org/10.1016/j.neuron.2017.09.007
Scott, RB, Samaha, J., Chrisley, R., and Dienes, Z., 2018. Prevailing theories of consciousness are challenged by novel cross-modal associations acquired between subliminal stimuli. Cognition, 175. pp. 169-185.
Sergent, C., Naccache, L., 2012. Imaging Neural Signatures of Consciousness: 'What', 'When', 'Where' and 'How' does it work? Arch Ital Biol., Jun-Sep;150(2-3):91-106. https://doi.org/10.4449/aib.v150i2.1270
Seth, A.K., Suzuki, K., Critchley, H.D., 2012. An Interoceptive Predictive Coding Model of Conscious Presence. Front. Psychol. 2. https://doi.org/10.3389/fpsyg.2011.00395
Silverstein, B.H., Snodgrass, M., Shevrin, H. Kushwaha, R., 2015. P3b, consciousness, and complex unconscious processing. Cortex 73, December 2015, pp. 216–227.
Spratling, M., 2016. A review of predictive coding algorithms. Brain Cogn. 112. https://doi.org/10.1016/j.bandc.2015.11.003
Sterzer, P., Adams, R.A., Fletcher, P., Frith, C., Lawrie, S.M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., Corlett, P.R., 2018. The Predictive Coding Account of Psychosis. Biol. Psychiatry, Mechanisms of Cognitive Impairment in Schizophrenia 84, 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015
Whyte, C.J., 2019. Integrating the Global Neuronal Workspace Into the Framework of Predictive Processing: Towards a Working Hypothesis. Conscious. Cogn. 73, 102763. https://doi.org/10.1016/j.concog.2019.102763
Wiese, W., Metzinger, T.K., 2017. Vanilla PP for Philosophers: A Primer on Predictive Processing. https://doi.org/10.15502/9783958573024
Wyart, V., Sergent, C., 2009. The Phase of Ongoing EEG Oscillations Uncovers the Fine Temporal Structure of Conscious Perception. The Journal of Neuroscience, October 14, 29(41):12839–12841. https://doi.org/10.1523/JNEUROSCI.3410-09.2009
Yon, D., Lange, F., Press, C., 2018. The Predictive Brain as a Stubborn Scientist. Trends Cogn. Sci. 23. https://doi.org/10.1016/j.tics.2018.10.003