frontiers in
PSYCHOLOGY

HYPOTHESIS AND THEORY ARTICLE
published: 02 April 2012
doi: 10.3389/fpsyg.2012.00096

Attention and conscious perception in the hypothesis testing brain

Jakob Hohwy*

Department of Philosophy, Monash University, Melbourne, VIC, Australia

Edited by:
Naotsugu Tsuchiya, Monash University, Australia

Reviewed by:
Andy Clark, University of Edinburgh, UK
Floris P. De Lange, Radboud University Nijmegen, Netherlands

****Correspondence:***
Jakob Hohwy, Department of Philosophy, Monash University, Clayton, VIC 3800, Australia.
e-mail: jakob.hohwy@monash.edu

Conscious perception and attention are difficult to study, partly because their relation to each other is not fully understood. Rather than conceiving and studying them in isolation from each other it may be useful to locate them in an independently motivated, general framework, from which a principled account of how they relate can then emerge. Accordingly, these mental phenomena are here reviewed through the prism of the increasingly influential predictive coding framework. On this framework, conscious perception can be seen as the upshot of prediction error minimization and attention as the optimization of precision expectations during such perceptual inference. This approach maps on well to a range of standard characteristics of conscious perception and attention, and can be used to interpret a range of empirical findings on their relation to each other.

Keywords: prediction error minimization, precision expectation, free energy, inattentional blindness, change blindness, unconscious processing

INTRODUCTION

The nature of attention is still unresolved, the nature of conscious perception is still a mystery – and their relation to each other is not clearly understood. Here, the relation between attention and conscious perception is reviewed through the prism of predictive coding. This is the idea that the brain is essentially a sophisticated hypothesis tester (Helmholtz, 1860; Gregory, 1980), which continually and at multiple spatiotemporal scales seeks to minimize the error between its predictions of sensory input and the actual incoming input (see Mumford, 1992; Friston, 2010). On this framework, attention and perception are two distinct, yet related aspects of the same fundamental prediction error minimization mechanism. The upshot of the review here is that together they determine which contents are selected for conscious presentation and which are not. This unifies a number of experimental findings and philosophical issues on attention and conscious perception, and puts them in a different light. The prediction error minimization framework transpires as an attractive, if yet still speculative, approach to attention and consciousness, and their relation to each other.

Attention is difficult to study because it is multifaceted and intertwined with conscious perception. Thus, attention can be endogenous (more indirect, top-down, or motivationally driven) or exogenous (bottom-up, attention grabbing); it can be focal or global; it can be directed at objects, properties, or spatial or temporal regions, and so on (Watzl, 2011a,b). Attentional change often seems accompanied by a change in conscious perception such that what grabs attention is a new stimulus, and such that whatever is attended to also populates consciousness. It can therefore be difficult to ascertain whether an experimental manipulation intervenes cleanly on attention or whether it intervenes on consciousness too (Van Boxtel et al., 2010).

Consciousness is difficult to study, partly because of the intertwinement with attention and partly because it is multifaceted too. Consciousness can apply to an overall state (e.g., awake vs. dreamless sleep) or a particular representation (e.g., conscious vs. unconscious processing of a face) all somehow tied together in the unity of the conscious stream (Bayne, 2010); it can pertain to the notion of a self (self-awareness) or just to being conscious (experience), and so on (Hohwy and Fox, 2012)[1]. There are widely accepted tools for identifying the neural correlates of conscious experience, though there is also some controversy about how cleanly they manipulate conscious states rather than a wide range of other cognitive processes (Hohwy, 2009). In the background is the perennial, metaphysical mind–body problem (Chalmers, 1996), which casts doubt on the possibility of ever achieving a fundamentally naturalist understanding of consciousness; (we will not discuss any metaphysics in this paper, however).

Functionally, attention is sometimes said to be an "analyzer," dissecting and selecting among the many possible and often competing percepts one has at any given time. Consciousness in contrast seems to be a "synthesizer," bringing together and organizing our multitudinous sensory input at any given time (Van Boxtel et al., 2010). On the other hand, attention may bring unity too, via binding (Treisman and Gelade, 1980), and consciousness also has a selective role when ambiguities in the sensory input are resolved in favor of one rather than the other interpretation, as seems to happen in binocular rivalry.

Attention and consciousness, then, are both difficult to define, to operationalize in functional terms, and to manipulate experimentally. Part of the trouble here has to do with the phenomena themselves, and possibly even their metaphysical underpinnings. But a large part of the trouble seems due to their intertwined relations. It is difficult to resolve these issues by appeal to

[1] In addition to perceptual forms of consciousness there is also a live debate, set aside here, about non-perceptual forms of consciousness, such as conceptual thought (Bayne and Montague, 2011).

commonsense or empirically informed conceptual analyses of each phenomenon in isolation of the other. For this reason it may be fruitful to appeal to a very general theoretical framework for overall brain function, such as the increasingly influential prediction error minimization approach, and review whether it implies coherently related phenomena with a reasonable fit to attention and conscious perception.

Section "Aspects of Prediction Error Minimization" describes heuristically the prediction error minimization approach. Section "Prediction Error and Precision" focuses on two aspects of this approach, here labeled accuracy and precision, and maps these onto perceptual inference and attention. Section "Conscious Perception and Attention as Determined by Precise Prediction Error Minimization" outlines why this mapping might be useful for understanding conscious perception and its relation to attention. In Section "Interpreting Empirical Findings in the Light of Attention as Precision Optimization," the statistical dimensions of precision and accuracy are used to offer interpretations of empirical studies of the relation between attention and consciousness. The final section briefly offers some broader perspectives.

ASPECTS OF PREDICTION ERROR MINIMIZATION

Two things motivate the idea of the hypothesis testing brain: casting a core task for the brain in terms of causal inference, and then appealing to the problem of induction.

The brain needs to represent the world so we can act meaningfully on it, that is, it has to figure out what in the world causes its sensory input. Representation is thereby a matter of causal inference. Causal inference however is problematic since a many–many relation holds between cause and effect: one cause can have many different effects, and one effect can have many different causes. This is the kernel of Hume's problem of induction (Hume, 1739–1740, Book I, Part III, Section vi): cause and effect are distinct existences and there are no necessary connections between distinct existences. Only with the precarious help of experience can the contingent links between them be revealed.

For the special case of the brain's attempt to represent the world, the problem of induction concerns how causal inference can be made "backwards" from the effects given in the sensory input to the causes in the world. This is the inverse problem, and it has a deep philosophical sting in the case of the brain. The brain *never* has independent access to both cause and effect because to have that it would already have had to solve the problem of representation. So it cannot learn from experience by just correlating occurrences of the two. It only has the effects to go by so must somehow begin the representational task *de novo*.

The prediction error minimization approach resolves this problem in time. The basic idea, described heuristically here, is simple whereas the computational details are complex (Friston, 2010). Sensory input is not just noise but has repeatable patterns. These patterns can give rise to expectations about subsequent input. The expectations can be compared to that subsequent input and the difference between them be measured. If there is a tight fit, then the pattern generating the expectation has captured a pattern in the real world reasonably well (i.e., the difference was close to expected levels of irreducible noise). If the fit is less good, that is, if there is a sizeable prediction error, then the states and parameters of the hypothesis or model of the world generating the expectation should be revised so that subsequent expectations will, over time, get closer to the actual input.

This idea can be summed up in the simple dictum that to resolve the inverse problem all that is needed is prediction error minimization. Expected statistical patterns are furnished by generative models of the world and instead of attempting the intractable task of inverting these models to extract causes from generated effects, prediction error minimization ensures that the model recapitulates the causal structure of the world and is implicitly inverted; providing a sufficient explanation for sensory input.

This is consistent with a Bayesian scheme for belief revision in the light of new evidence, and indeed both Bayes as well as Laplace (before he founded classical frequentist statistics) developed their theories in response to the Humean-inspired inverse problem (McGrayne, 2011). The idea is to weight credence in an existing model of the world by how tightly it fits the evidence (i.e., the likelihood or how well it predicts the input) as well as how likely the model is in the first place (i.e., the prior probability or what the credence for the model was before the evidence came in).

The inverse problem is then resolved because, even though there is a many–many relation between causes in the world and sensory effects, some of the relations are weighted more than others in an optimally Bayesian way. The problem is solved *de novo*, without presupposing prior representational capability, because the system is supervised not by another agent, nor by itself, but by the very statistical regularities in the world it is trying to represent.

This key idea is then embellished in a number of different ways, all of which have bearing on attention and conscious perception.

HIERARCHY

The prediction error minimization mechanism sketched above is a general type of statistical building block that is repeated throughout levels of the cortical hierarchy such that there is recurrent message passing between levels (Mumford, 1992). The input to the system from the senses is conceived as prediction error and what cannot be predicted at one level is passed on to the next. In general, low levels of the hierarchy predict basic sensory attributes and causal regularities at very fast, millisecond, time scales, and more complex regularities, at increasingly slower time scales, are dealt with at higher levels (Friston, 2008; Kiebel et al., 2008, 2010; Harrison et al., 2011). Prediction error is concurrently minimized across all levels of the hierarchy, and this unearths the states and parameters that represent the causal structure and depth of the world.

CONTEXTUAL PROBABILITIES

Predictions at any level are subject to contextual modulation. This can be via lateral connectivity, that is, by predictions or hypotheses at the same hierarchical level, or it can be through higher level control parameters shaping low level predictions by taking slower time scale regularities into consideration. For example, the low level dynamics of birdsong is controlled by parameters from higher up pertaining to slower regularities about the size and strength of the bird doing the singing (Kiebel et al., 2010). Similarly, it may be that the role of gist perception is to provide contextual clues for fast classification of objects in a scene (Kveraga et al., 2007). The entire cortical hierarchy thus recapitulates the causal structure of

the world, and the bigger the hierarchy the deeper the represented causal structure.

EMPIRICAL BAYES

For any appeal to Bayes, the question arises where do the priors come from (Kersten et al., 2004)? One scheme for answering this, and evading charges of excessive subjectivity, is empirical Bayes where priors are extracted from hierarchical statistical learning (see, e.g., Casella, 1992). In the predictive coding scheme this does not mean going beyond Bayes to frequentism. (Empirical) Priors are sourced from higher levels in the hierarchy, assuming they are learned in an optimally Bayesian fashion (Friston, 2005). The notion of hierarchical inference is crucial here, and enables the brain to optimize its prior beliefs on a moment to moment basis. Many of these priors would be formed through long-term exposure to sensory contingencies through a creature's existence but it is also likely that some priors are more hard-wired and instantiated over an evolutionary time-scale; different priors should therefore be malleable to different extents by the creature's sensation.

FREE ENERGY

In its most general formulation, prediction error minimization is a special case of free energy minimization, where free energy (the sum of squared prediction error) is a bound on information theoretical surprise (Friston and Stephan, 2007). The free energy formulation is important because it enables expansion of the ideas discussed above to a number of different areas (Friston, 2010). Here, it is mainly the relation to prediction error minimization that will be of concern. Minimizing free energy minimizes prediction error and implicitly surprise. The idea here is that the organism cannot directly minimize surprise. This is because there is an infinite number of ways in which the organism could seek to minimize surprise and it would be impossibly expensive to try them out. Instead, the organism can test predictions against the input from the world and adjust its predictions until errors are suppressed. Even if the organism does not know what will surprise it, it can minimize the divergence between its expectations and the actual inputs encountered. A frequent objection to the framework is that prediction error and free energy more generally can be minimized by committing suicide since nothing surprises a dead organism. The response is that the moment an organism dies it experiences a massive increase in free energy, as it decomposes and is unable to predict anything (there is more to say on this issue, see Friston et al., in press; there is also a substantial issue surrounding how these types of ideas can be reconciled with evolutionary ideas of survival and reproduction, for discussion see, Badcock, 2012).

ACTIVE INFERENCE

A system without agency cannot minimize surprise but only optimize its models of the world by revising those models to create a tight free energy bound on surprise. To minimize the surprise it needs to predict how the system's own intervention in the world (e.g., movement) could change the actual input such as to minimize free energy. Agency, in this framework, is a matter of selectively sampling the world to ensure prediction error minimization across all levels of the cortical prediction hierarchy (Friston et al., 2009, 2011). To take a toy example: an agent sees a new object such as a bicycle, the bound on this new sensory surprise is minimized, and the ensuing favored model of the world lets the agent predict how the prediction error landscape will change given his or her intervention (e.g., when walking around the bike). This prediction gives rise to a prediction error that is not minimized until the agent finds him or herself walking around the bike, hence the label "active inference." If the initial model was wrong, then active inference fails to be this kind of self-fulfilling prophecy (e.g., it was a cardboard poster of a bike). Depending on the depth of the represented causal hierarchy this can give rise to very structured behavior (e.g., not eating all your food now even though you are hungry and instead keeping some for winter, based on the prediction this will better minimize free energy).

There is an intuitive seesaw dynamic here between minimizing the bound and actively sampling the world. It would be difficult to predict efficiently what kind of sampling would minimize surprise if the starting point was a very poor, inaccurate, bound on surprise. Similarly, insofar as selective sampling never perfectly minimizes surprise, new aspects of the world are revealed, which should lead to revisiting the bound on surprise. It thus pays for the system to maintain both perceptual and active inference.

TOP-DOWN AND BOTTOM-UP

This framework comes with a re-conceptualization of the functional roles of the bottom-up driving signal from the senses, and the top-down or backward modulatory signal from higher levels. The bottom-up signal is not sensory information *per se* but instead just prediction error. The backward signal embodies the causal model of the world and the bottom-up prediction error is then essentially the supervisory feedback on the model (Friston, 2005). It is in this way the sensory input ensures the system is supervised, not by someone else nor by itself, but by the statistical regularities of the world.

The upshot is an elegant framework, which is primarily motivated by principled, philosophical and computational concerns about representation and causal inference. It is embellished in a number of ways that capture many aspects of sensory processing such as context-dependence, the role of prior expectations, the way perceptual states comprise sensory attributes at different spatiotemporal resolutions, and even agency. We shall appeal to all these elements as predictive coding is applied to attention and conscious perception.

PREDICTION ERROR AND PRECISION

As discussed above, there are two related ways that prediction error can be minimized: either by changing the internal, generative model's states, and parameters in the light of prediction error, or keeping the model constant and selectively sampling the world and thereby changing the input. Both ways enable the model to have what we shall here call *accuracy*: the more prediction error is minimized, the more the causal structure of the world is represented[2].

[2]There is a simplification here: surprise has both accuracy and complexity components, such that minimizing surprise or free energy increases accuracy while minimizing complexity. This ensures the explanations for sensory input are parsimonious and will generalize to new situations; c.f., Occam's razor.

So far, this story leaves out a crucial aspect of perceptual inference concerning *variability* of the prediction error. Prediction error minimization of the two types just mentioned assumes noise to be constant, and the variability of all prediction errors therefore the same. This assumption does not actually hold as noise or uncertainty is state dependent. Prediction error that is unreliable due to varying levels of noise in the states of the world is not a learning signal that will facilitate confident veridical revision of generative models or make it likely that selective sampling of the world is efficient. Prediction error minimization must therefore take variability in prediction error messaging into consideration – it needs to assess the precision of the prediction error.

Predictions are tested in sensory sampling: given the generative model a certain input is predicted where this input can be conceived as a distribution of sensory samples. If the actual distribution is different from the expected distribution, then a prediction error is generated. One way to assess a difference in distributions is to assess central tendency such as the mean. However, as is standard in statistical hypothesis testing, even if the means seem different (or not) the variability may preclude a confident conclusion that the two distributions are different (or not). Hence, any judgment of difference must be weighed by the magnitude of the variability – this is a requirement for trusting prediction error minimization.

The inverse of variability is the precision (inverse dispersion or variance) of the distribution. In terms of the framework used here, when the system "decides" whether to revise internal models in the light of prediction errors and to sample the world accordingly, those errors are weighted by their precisions. For example, a very imprecise (i.e., noisy, variable) prediction error should not lead to revision, since it is more likely to be a random upshot of noise for a given sensory attribute.

However, the rule cannot be simply that the more the precision the stronger the weight of the prediction error. Our expectations of precision are context dependent. For example, precisions in different sensory modalities differ (for an example, see Bays and Wolpert, 2007), and differ within the same modality in different contexts and for different sensory attributes. Sometimes it may be that one relatively broad, imprecise distribution should be weighed more than another narrower, precise distribution. Similarly, an unusually precise prediction error may be highly inaccurate as a result of under-sampling, for example, and should not lead to revision. In general, the precision weighting should depend on prior learning of regularities in the actual levels of noise in the states of the world and the system itself (e.g., learning leading to internal representations of the regularity that sensory precision tends to decline at dusk).

There is then a (second order) perceptual inference problem because the magnitude of precision cannot be measured absolutely. It must be assessed in the light of *precision expectations*. The consequence is that generative models must somehow embody expectations for the precision of prediction error, in a context dependent fashion. Crucially, the precision afforded a prediction has to be represented; in other words, one has to represent the known unknowns.

If precision expectations are optimized then prediction error is weighted accurately and replicates the precisions in the world. In terms of perceptual inference, the learning signal from the world will have more weight from units expecting precision, whereas top-down expectations will have more influence on perception when processing concerns units expecting a lot of imprecision; one's preconceptions play a bigger role in making sense of the world when the signal is deemed imprecise (Hesselmann et al., 2010). This precision processing is thought to occur in synaptic error processing such that units that expect precision will have more weight (synaptic gain) than units expecting imprecision (Friston, 2009).

Given a noisy world and non-linear interactions in sensory input, first order statistics (prediction errors) and second order statistics (the precision of prediction errors) are then necessary and jointly sufficient for resolving the inverse problem. In what follows, the optimization of representations is considered in terms of both *precision* and *accuracy*, precision refers to the inverse amplitude of random fluctuations around, or uncertainty about, predictions; while accuracy (with a slight abuse of terminology) will refer to the inverse amplitude of prediction errors *per se*. Minimizing free energy or surprise implies the minimization of precise prediction errors; in other words, the minimization of the sum of squared prediction error and an optimal estimate of precision.

Using the terminology of accuracy and precision is useful because it suggests how the phenomena can come apart in a way that will help in the interpretation of the relation between consciousness and attention. It is a trivial point that precision and accuracy can come apart: a measurement can be accurate but imprecise, as in feeling the child's fever with a hand on the forehead or it can be very precise but inaccurate, as when using an ill calibrated thermometer. This yields two broad dimensions for perceptual inference in terms of predictive coding: accuracy (via expectation of sensory input) and precision (via expectation of variability of sensory input). These can also come apart. Some of the states and parameters of an internal model can be inaccurate and yet precise (being confident that the sound comes from in front of you when it really comes from behind, Jack and Thurlow, 1973). Or they can be accurate and yet, imprecise (correctly detecting a faint sound but being uncertain about what to conclude given a noisy background).

With this in mind, assume now that *conscious perception* is determined by the prediction or hypothesis with the highest overall posterior probability – which is overall best at minimizing prediction error (this assumption is given support in the next section). That is, conscious perception is determined by the strongest "attractor" in the free energy landscape; where, generally speaking, greater precision leads to higher conditional confidence about the estimate and a deeper, more pronounced minimum in the free energy landscape.

On this assumption, precision expectations play a key role for conscious perception. We next note the proposal, which will occupy us in much of the following, that optimization of precision expectations maps on to attention (Friston, 2009). It is this mapping that will give substance to our understanding of the relation between attention and consciousness. It is a promising approach because precision processing, in virtue of its relation to accuracy, has the kind of complex relation to prediction error minimization that seems appropriate for capturing both the commonsense notion that conscious perception and attention are intertwined

and also the notion that they are separate mechanisms (Koch and Tsuchiya, 2007; Van Boxtel et al., 2010).

We can usefully think of this in terms of a system such that, depending on context (including experimental paradigms in the lab), sensory estimates may be relatively accurate and precise, inaccurate and imprecise, accurate and imprecise, or inaccurate and precise. With various simplifications and assumptions, this framework can then be sketched as in **Figure 1**.

By and large, conscious perception will be found for states that are both accurate and precise but may also be found for states that are relatively accurate and yet imprecise, and *vice versa*. Two or more competing internal models or hypotheses about the world can have different constellations of precision and accuracy: a relatively inaccurate but precise model might determine conscious perception over a competing accurate but imprecise model, and *vice versa*. Similarly, a state can evolve in different ways: it can for example begin by being very inaccurate and imprecise, and thus not determining conscious perception but attention can raise its conditional confidence and ensure it does get to determine conscious content.

On this framework, it should then also be possible to speak to some of the empirical findings of dissociations between attention and consciousness. A case of attention without consciousness would be where precision expectations are high for a state but prediction error for it is not well minimized (expecting a precise signal, or, expecting inference to be relatively bottom-up driven). A case of consciousness without attention would be where prediction error is well minimized but where precision is relatively low (expecting signals to be variable, or, expecting inference to be relatively top-down driven). It is difficult to say precisely what such states would be like. For example, a conscious, inattentive state might have a noisy, fuzzy profile, such as gist perception may have (Bar, 2007). It is also possible that increased reliance on top-down, prior beliefs could in fact paradoxically sharpen the representational profile (Ross and Burr, 2008)[3]. In general, in both types of cases, the outcome would be highly sensitive to the context of the overall free energy landscape, that is, to competing hypotheses and their precision expectations.

Section "Interpreting Empirical Findings in the Light of Attention as Precision Optimization" will begin the task of interpreting some studies in the field according to these accuracy and precision dimensions. The next section, however, will provide some *prima facie* motivation for this overall framework.

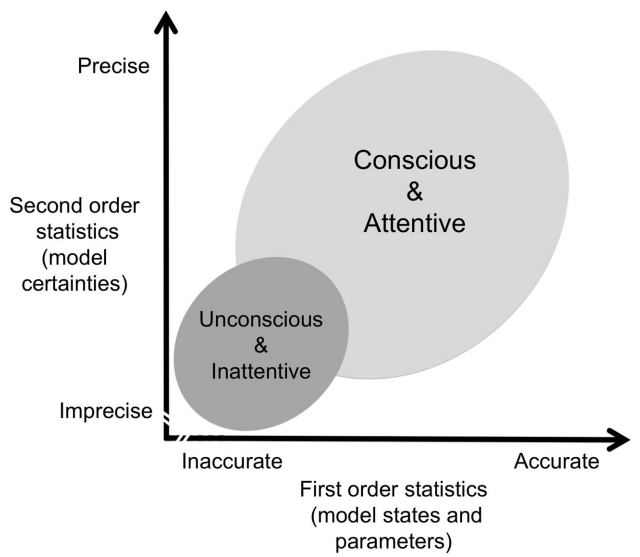

FIGURE 1 | Schematic of statistical dimensions of conscious perception. The accuracy afforded by first order statistics refers to the inverse amplitude of prediction errors *per se*, while the precision afforded by second order statistics refers to the inverse amplitude of random fluctuations around, or uncertainty about, predictions. This allows for a variety of different types of states such that in general, and depending on context, inattentive but conscious states would cluster towards the lower right corner and attentive but unconscious states would cluster towards the upper left; see main text for further discussion.

CONSCIOUS PERCEPTION AND ATTENTION AS DETERMINED BY PRECISE PREDICTION ERROR MINIMIZATION

In this section, conscious perception and attention are dealt with through the prism of predictive coding. Though the evidence in favor of this approach is growing (see the excellent discussion in Summerfield and Egner, 2009) much of this is still speculative[4]. The core idea is that conscious perception correlates with activity, spanning multiple levels of the cortical hierarchy, which best suppresses precise prediction error: what gets selected for conscious perception is the hypothesis or model that, given the widest context, is currently most closely guided by the current (precise) prediction errors[5].

Conscious perception can then thought to be at the service of representing the world, and the currently best internal, generative model is the one that most probably represents the causal structure of the world. Predictions by other models may also be able to suppress prediction error, but less well, so they are not selected. Conversely, often some other, possible models could be even better at suppressing prediction error but if the system has not learnt them yet, or cannot learn them, it must make do with the best model it has.

It follows that the predictions of the currently best model can actually be rather inaccurate. However, if it has no better competitor then it will win and get selected for consciousness. Conscious

[3]There is also a very good question here about how this kind of confidence assessment fits with the psychological confidence of the organism, which appears a defining feature of consciousness, and which is often assessed in confidence ratings. (Thanks to a reviewer for raising this issue).

[4]A further disclaimer: the speculation that conscious perception is a product of accuracy and precision in predictive coding is a limited speculation about an information processing mechanism. It is not a speculation about *why* experience is conscious rather than not conscious – predictive coding can after all be implemented in unconscious machines. The mystery of consciousness will remain untouched.

[5]This claim depends on optimal Bayesian inference actually being able to recapitulate the causal structure of the world. Here we bracket for philosophical debate the fact that this assumption breaks down for perfect skeptical scenarios, such as Cartesian deceiving demons or evil scientists manipulating brains in vats, where minimizing free energy does not reveal the true nature of the world. We also bracket deeper versions of the problem of induction, such as the new riddle of induction (Goodman, 1955). though we note that when two hypotheses are equally good at predicting new input the free energy principle prefers the one with the smallest complexity cost.

perception can then be far from veridical, in spite of its representational nature. This makes room for an account of illusory and hallucinatory perceptual content, which is an important desideratum on accounts of conscious perception. These would be cases where, for different reasons, poor models are best at precisely explaining away incoming data only because their competitors are even poorer.

The job of the predictive coding system is to attenuate sensory input by treating it as information theoretical surprise and predicting it as perfectly as possible. As the surprise is attenuated, models should stop being revised and predictive activity progressively cease throughout the hierarchy. This seems consistent with repetition suppression (Grill-Spector et al., 2006) where neural activity ceases in response to expected input in a manner consistent with prediction error minimization (Summerfield et al., 2008; Todorovic et al., 2011). At the limit it should have consequences for conscious perception too. 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. Presumably there is almost always some input to some consumer systems in the brain (including during dreaming) but conceivably something close to this happens when stabilized retinal images fade from consciousness (Ditchburn and Ginsborg, 1952). Because such stimuli move with eye and head movement predictive exploration of them is quickly exhausted.

Conscious perception is often rich in sensory attributes, which are neatly bound together even though they are processed in a distributed manner throughout the brain. The predictive coding account offers a novel approach to this "binding" aspect of conscious perception. Distributed sensory attributes are bound together by the causal inference embodied in the parameters of the generative model. The model assumes, for example, that there is a red ball out there so will predict that the redness and the bouncing object co-occur spatiotemporally. The binding problem (Treisman, 1996) is then dealt with by default: the system does not have to operate in a bottom-up fashion and first process individual attributes and then bind them. Instead, it assumes bound attributes and then predicts them down through the cortical hierarchy. If they are actually bound in the states of the world, then this will minimize prediction error, and they will be experienced as such.

It is a nice question here what it means for the model with the highest posterior probability to be "selected for consciousness." We can only speculate about an answer but it appears that on the predictive coding framework there does not have to be a specific selection mechanism (no "threshold" module, cf. Dennett, 1991). When a specific model is the one determining the consciously perceived content it is just because it best minimizes prediction error across most levels of the cortical hierarchy – it best represents the world given all the evidence and the widest possible context. This is the model that should be used to selectively sample the world to minimize surprise in active inference. Competing but less probable models cannot simultaneously determine the target of active inference: the models would be at cross-purposes such that the system would predict more surprise than if it relies on one model alone (for more on the relation between attention and action, see Wu, 2011).

Though there remain aspects of consciousness that seem difficult to explain, such as the conscious content of imagery and dreaming, this overall approach to conscious perception does then promise to account for a number of key aspects of consciousness. The case being built here is mainly theoretical. There is not yet much empirical evidence for this link to conscious perception, though a recent dynamical causal modeling study from research in disorders of consciousness (vegetative states and minimally conscious states) suggests that what is required for an individual to be in an overall conscious state is for them to have intact connectivity consistent with predictive coding (Boly et al., 2011).

As we saw earlier, in the normal course of events, the system is helped in this prediction error minimization task by precision processing, which (following Feldman and Friston, 2010) was claimed to map on to attention such that attention is precision optimization in hierarchical perceptual inference. A prediction error signal will have a certain absolute dispersion but whether the system treats this as precise or not depends on its precision expectations, which may differ depending on context and beliefs about prior precision. Precise prediction errors are reliable signals and therefore, as described earlier, enable a more efficient revision of the model in question (i.e., a tighter bound and better active inference). If that model then, partly resulting from precision optimization, achieves the highest posterior probability, it will determine the content of conscious perception. This begins to capture the functional role often ascribed to attention of being a gating or gain mechanism that somehow optimizes sensory processing (Hillyard and Mangun, 1987; Martinez-Trujillo and Treue, 2004). As shall be argued now, it can reasonably account for a wider range of characteristics of attention.

EXOGENOUS ATTENTION

Stimuli with large spatial contrast and/or temporal contrast (abrupt onset) tend to "grab" attention bottom-up, or exogenously. These are situations where there is a relatively high level of sensory input, that is, a stronger signal. Given an expectation that stronger signals have better signal to noise ratio (better precision), than weaker signals (Feldman and Friston, 2010, p. 9; Appendix), error units exposed to such signals should thus expect high precision and be given larger gain. As a result, more precise prediction error can be suppressed by the model predicting this new input, which is then more likely to be the overall winner populating conscious experience. Notice that this account does not mention expectations about *what* the signal stems from, only about the signal's reliability. Also notice that this account does not guarantee that what has the highest signal to noise ratio will end up populating consciousness, it may well be that other models have higher overall confidence or posterior probability.

ENDOGENOUS ATTENTION

Endogenous attention is driven more indirectly by probabilistic context. Beginning with endogenous cueing, a central cue pointing left is itself represented with high precision prediction error (it grabs attention) and in the parameters of the generative model

this cue representation is related to the representation of a stimulus to the left, via a learned causal link. This reduces uncertainty about what to predict there (increases prior probability for a left target) and it induces an expectation of high precision for that region. When the stimulus arrives, the resulting gain on the error units together with the higher prior help drive a higher conditional confidence for it, making it likely it is quickly selected for conscious perception.

The idea behind endogenous attention is then that it works as an increase in baseline activity of neuronal units encoding beliefs about precision. There is evidence that such increase in activity prior to stimulus onset is specific to precision expectations. The narrow distributions associated with precise processing tell us that in detection tasks the precision-weighted system should tend to respond when and only when the target appears. And indeed such baseline increases do bias performance in favor of hits and correct rejections (Hesselmann et al., 2010). In contrast, if increased baseline activity had instead been a matter of mere accumulation of evidence for a specific stimulus (if it had been about accuracy and not precision), then the baseline increase should instead have biased toward hits and false alarms.

A recent paper directly supports the role of endogenous attention as precision weighting (Kok et al., 2011). As we have seen, without attention, the better a stimulus is predicted the more attenuated its associated signal should be. Attention should reverse this attenuation because it strengthens the prediction error. However, attention depends on the predictability of the stimulus: there should be no strong expectation that an unpredicted stimulus is going to be precise. So there should be less attention-induced enhancement of the prediction error for unpredicted stimuli than for better predicted stimuli. Using fMRI, Kok et al. very elegantly provides evidence for this interaction in early visual cortex (V1).

In more traditional cases of endogenous attention (e.g., the individual deciding herself to attend left) the cue can be conceived as a desired state, for example, that something valuable will be spotted to the left. This would then generate an expectation of precision for that region such that stimuli located there are more likely to be detected. Endogenous attention of this sort has a volitional aspect: the individual decides to attend and acts on this decision. Such agency can range from sensorimotor interaction and experimentation to a simple decision to fixate on something. This agential aspect suggests that part of attention should belong with active inference (selective sampling to minimize surprise). The idea here would be that the sampling is itself subject to precision weighting. This makes sense since the system will not know if its sampling satisfies expectations unless it can assess the variability in the sampling. Without such an assessment, the system will not know whether to keep sampling on the basis of a given model or whether the bound on the model itself needs to be re-assessed. In support of this, there is emerging evidence that precision expectations are also involved in motor behavior (Brown et al., 2011).

BIASED COMPETITION

An elegant approach to attention begins with the observation that neurons respond optimally to one object or property in their receptive field so that if more than one object is present, activity decreases unless competition between them is resolved. The thought is that attention can do this job, by biasing one interpretation over another (Desimone and Duncan, 1995). Attention is thus required to resolve ambiguities of causal inference incurred by the spatial architecture of the system. Accordingly, electrophysiological studies show decreased activity when two different objects are present in a neuron's receptive field, and return to normal levels of activity when attention is directed toward one of them (Desimone, 1998).

The predictive coding framework augmented with precision expectations should be able to encompass biased competition. This is because, as mentioned, precision can modulate perceptual inference when there are two or more competing, and perhaps equally accurate, models. Indeed, computational simulation shows precision-weighted predictive coding can play such a biasing role in a competitive version of the Posner paradigm where attention is directed to a cued peripheral stimulus rather than a competing non-cued stimulus. A central cue thus provides a context for the model containing the cued stimulus as a hidden cause. This drives a high precision expectation for that location, which ensures relatively large gain, and quicker response times, when those error units are stimulated. This computational model nicely replicates psychophysics and electrophysiological findings (Feldman and Friston, 2010, pp. 14–15).

Attentional competition is then not a matter somehow of intrinsically limited processing resources or of explicit competition. It is a matter of optimal Bayesian inference where only one model of the causal regularities in the world can best explain away the incoming signal, given prior learning, and expectations of state-dependent levels of noise.

Binding of sensory attributes by a cognitive system was mooted above as a natural element of predictive coding. Attention is also thought to play a role for binding (Treisman and Gelade, 1980; Treisman, 1998) perhaps via gamma activity (Treisman, 1999) such that synchronized neurons are given greater gain. Again, this can be cast in terms of precision expectations: sensory attributes bound to the same object are mutually predictive and so if the precision-weighted gain for one is increased it should increase for the other too. Though this is speculative, the predictive coding framework could here elucidate the functional role of increased gamma activity and help us understand how playing this role connects to attention and conscious perception.

Perhaps we should pause briefly and ask why we should adopt this framework for attention in particular – what does it add to our understanding of attention to cast it in terms of precision expectations? A worry could be that it is more or less a trivial reformulation of notions of gain, gating, and bias, which has long been used to explicate attention in a more or less aprioristic manner. The immediate answer is that this account of attention goes beyond mere reformulations of known theories, not just because its basic element is precision, but also because it turns on learning precision regularities in the world so different contexts will elicit different precision expectations. This is crucial because optimization of precision is context dependent and thus requires appeal to just the kind of predictive framework used here.

There is also a more philosophical motivation for adopting this approach. Normally, an account of attention would begin with

some kind of operational, conceptual analysis of the phenomenon: attention has to do with salience, with some kind of selection of sensory channels, resource limitations, and so on. Then the evidence is consulted and theories formulated about neural mechanisms that could underpin salience and selection etc. This is a standard and fruitful approach in science. But sometimes taking a much broader approach gives a better understanding of the nature of the phenomenon of interest and its relation to other phenomena (*cf.* explanation by unification, Kitcher, 1989). In our case, a very general conception of the fundamental computational task for the brain defines two functional roles that must be played: estimation of states and parameters, and estimation of precisions. Without beginning from a conceptual analysis of attention, we then discover that the element of precision processing maps on well to the functional role we associate with attention. This discovery tells us something new about the nature of attention: the reason why salience and selection of sensory channels matter, and the reason why there appears to be resource limitations on attention, is that the system as such must assess precisions of sensory estimates and weight them against each other.

Viewing attention from the independent vantage point of the requirements of predictive coding also allows us to revise the concept of attention somewhat, which can often be fruitful. For example, there is no special reason why attention should always have to do with conscious perception, given the ways precision and accuracy can come apart; that is, there may well be precision processing – attention – outside consciousness. The approach suggests a new way for us to understand how attention and perception can rely on separate but related mechanisms. This is the kind of issue to which we now turn.

INTERPRETING EMPIRICAL FINDINGS IN THE LIGHT OF ATTENTION AS PRECISION OPTIMIZATION

The framework for conscious perception sketched in Section "Prediction Error and Precision" (see **Figure 1**) implied that studies of the relation between consciousness and attention can be located according to the dimensions of accuracy and precision. We now explore if this implication can reasonably be said to hold for a set of key findings concerning: inattentional blindness, change blindness, the effects of short term and sustained covert attention on conscious perception, and attention to unconscious stimuli.

The tools for interpreting the relevant studies must be guided by the properties of predictive coding framework we have set out above, so here we briefly recapitulate: (1) even though accuracy and precision are both necessary for conscious perception, it does not follow that the single *most* precise or the *most* accurate estimate in a competing field of estimates will populate consciousness: that is determined by the overall free energy landscape. For example, it is possible for the highest overall posterior probability to be determined by an estimate having high accuracy and relatively low precision even if there is another model available that has relatively low accuracy yet high precision, and so on. (2) Attention in the shape of precision expectation modulates prediction error minimization subject to precisions predicted by the context, including cues and competing stimuli; it can do this for prediction errors of different accuracies. (3) Precision weighting only makes sense if weights sum to one so that as one goes up the others must go down. Similarly, as the probability of one model goes up the probability of other models should go down – the other models are explained away if one model is able to account for and suppress the sensory input. This gives rise to model competition. (4) Conscious experience of unchanging, very stable stimuli will tend to be suppressed over time, as prediction error is explained away and no new error arises. (5) Agency is active inference: a model of the agent's interaction with the world is used to selectively sample the world such as to minimize surprise. This also holds for volitional aspects of attention, such as the agency involved in endogenous attention to a spatial location.

The aim now is to use these properties of predictive coding to provide a coherent interpretation of the set of very different findings on attention and consciousness.

TYPES OF INATTENTIONAL BLINDNESS

The context for a stimulus can be a cue or an instruction or other sensory information, or perhaps a decision to attend. Various elements of this context can give a specific generative model two advantages: it can increase priors for its states and parameters (for this part of the view, see also Rao, 2005) and it can bias selection of that model via precision weighting. When the target stimulus comes, attention has thus already given the model for that stimulus a probabilistic advantage. If in contrast the context is invalid (non-predictive) and a different target stimulus occurs, the starting point for the model predicting it can be much lower both in terms of prior probability and in terms of precision expectation. If this lower starting point is sufficiently low, and if the invalidly contextualized stimulus is not itself strongly attention grabbing (is not abrupt in some feature space such as having sharp contrast or temporal onset), then "the invalid target may never actually be perceived" (Feldman and Friston, 2010, pp. 9–10).

This is then what could describe forms of inattentional blindness where an otherwise visible stimulus is made invisible by attending to something at a different location: an attentional task helps bias one generative model over models for unexpected background or peripheral stimuli. A very demanding attentional task would have very strong bias from precision weighting, and correspondingly the weight given to other models must be weakened. This could drive overall posterior probability below selection for consciousness, such that not even the gist of, for example, briefly presented natural scenes is perceived.

It is natural to conclude in such experiments that attention is a necessary condition for conscious perception since unattended stimuli are not seen, and as soon as they are seen performance on the central task decreases (Cohen et al., 2011). This is right in the sense that any weighting of precision to the peripheral or background stimulus must go with decreased weight to the central task. However, the more fundamental truth here is that in a noisy world precision weighting is necessary for conscious perception so that at the limit, where noise expectations are uniform, there could be conscious perception even though attention plays very little actual role.

When inattentional blindness is less complete, the gist of briefly presented natural scenes can be perceived (see, Van Boxtel et al., 2010). This is consistent with relatively low precision expectation since gist is by definition imprecise. So in this case some, but

relatively little prediction error is allowed through for the natural scene, leaving only little prediction error to explain away. It seems likely that this could give rise to gist rather than full perception. However, the distinction between gist and full perception is not well understood and there are more specific views on gist perception, also within the broad predictive coding framework (Bar, 2003).

In some cases of inattentional blindness, large and otherwise very salient stimuli can go unnoticed. Famously, when counting basketball passes a gorilla can be unseen, and when chasing someone a nearby fistfight can be unseen (Simons and Chabris, 1999; Chabris et al., 2011). This is somewhat difficult to explain because endogenous attention as described so far should raise the baseline for precision expectation for a specific location such that any stimulus there, whether it is a basketball pass or a gorilla, should be more likely to be perceived. A smaller proportion of participants experience this effect, so it does in fact seem harder to induce blindness in this kind of paradigm than paradigms using central–peripheral or foreground-background tasks. For those who do have inattentional blindness under these conditions, the explanation could be high precision expectations for the basketball passes specifically, given the context of the passes that have occurred before the gorilla enters. This combines with the way this precision error has driven up the conditional confidence of the basketball model, explaining away the gorilla model, even if the latter is fed some prediction error. This more speculative account predicts that inattentional blindness should diminish if the gorilla, for example, occurs at the beginning of the counting task.

This is then a way to begin conceptualizing feature- and object-based based attention instead of purely spatial attention. Van Boxtel et al. (2010) suggest that in gorilla type cases the context provided by the overall scene delivers a strong gist that overrides changes that fit poorly with it: "subjects do perceive the gist of the image correctly, interfering with detection of a less meaningful change in the scene as if it was filled in by the gist." The predictive coding approach can offer an explanation of this kind of interference in probabilistic terms.

A further aspect can be added to this account of inattentional blindness. Attending, especially endogenous attending, is an activity. As such, performing an attention demanding task is a matter of active inference where a model of the world is used to selectively sample sensory input to minimize surprise. This means that high precision input are expected and sampled on the basis of one, initial (e.g., "basketball") model, leaving unexpected input such as the occurrence of a gorilla with low weighting. Since the active inference required to comply with an attentional task must favor one model in a sustained way, blindness to unexpected stimuli follows.

The benefit of sustained attention viewed as active inference is then that surprise can be minimized with great precision, given an initial model's states and parameters. On the other hand, the cost of sustained attention is that the prediction error landscape may change during the task; increasing the free energy and making things evade consciousness.

It can thus be disadvantageous for a system to be stuck in active inference and neglecting to revisit the bound on surprise by updating the model (e.g., if the gorilla is real and angry). Perhaps the reason attention can be hard to maintain is that to avoid such disadvantage the system continually seeks, perhaps via spontaneous fluctuations, to alternate between perceptual and active inference. Minor lapses of attention (e.g., missing a pass) could thus lead to some model revision and conscious perception; if the model revision has relatively low precision it may just give rise to gist perception (e.g., "some black creature was there").

It is interesting here to speculate further that the functional role of exogenous attention can be to not only facilitate processing of salient stimuli but in particular to make the system snap out of active inference, which is often associated with endogenous attention, and back into revision of its generative model. Exogenous and endogenous attention seem to have opposing functional roles in precision optimization.

There remains the rather important and difficult question whether or not the unseen stimulus is in fact consciously perceived but not accessible for introspective report, or whether it is not consciously perceived at all; this question relates to the influential distinction between access consciousness and phenomenal consciousness (Block, 1995, 2008). To some, this question borders on the incomprehensible or at least untestable (Cohen and Dennett, 2011), and there is agreement it cannot be answered directly (e.g., by asking participants to report). Instead some indirect, abductive answer must be sought. We cannot answer this question here but we can speculate that the common intuition that there is both access *and* phenomenal consciousness is fueled by the moments of predictive coding such that (i) access consciousness goes with active inference (i.e., minimizing surprise though agency, which requires making model parameters and states available to control systems), and (ii) phenomenal consciousness goes with perceptual inference (i.e., minimizing the bound on surprise by more passively updating model parameters and states).

If this is right, then a prediction is that in passive viewing, where attention and active inference is kept as minimal as possible, there should be more possibility of having incompatible conscious percepts at the same time, since without active inference there is less imperative to favor just one initial model. There is some evidence for this in binocular rivalry where the absence of attention seems to favor fusion (Zhang et al., 2011).

Overall, some inroads on inattentional blindness can be made by an appeal to precision expectations giving the attended stimulus a probabilistic advantage. A more full, and speculative, explanation conceives attention in agential terms and appeals to the way active inference can lead to very precise but eventually overall inaccurate perceptual states.

CHANGE BLINDNESS

These are cases where abrupt and scene-incongruent changes like sudden mudsplashes attract attention and make invisible other abrupt but scene-congruent changes like a rock turning into a log or an aircraft engine going missing (Rensink et al., 1997). Only with attention directed at (or on repeated exposures grabbed by) the scene-congruent change will it be detected. This makes sense if the distractor (e.g., mudsplashes) has higher signal strength than the masked stimuli because, as we saw, there is a higher precision expectation for stronger signals. This weights prediction error for a mudsplash model rather than for a natural scenery model with logs or aircrafts. Even if both models are updated in the light of their

respective prediction errors from the mudsplashes and the rock changing to the log, the mudsplash model will have higher conditional confidence because it can explain away precisely a larger part of the bottom-up error signal.

More subtly, change blindness through attention grabbing seems to require that the abrupt stimuli activate a competing model of the causes in the world. This means that the prediction error can be relevant to the states and parameters of one of these models. Thus, the mudsplashes mostly appear to be superimposed on the original image, which activates a model with parameters for causal interaction between mudsplashes and something like a static photo. In other words, the best explanation for the visual input is the transient occlusion or change to a photo, where, crucially, we have strong prior beliefs that photographs do not change over short periods of time. This contrasts with the situation prior to the mudsplashes occurring where the model would be tuned more to the causal relations inherent in the scene itself (that is, the entire scene is not treated as a unitary object that can be mudsplashed). With two models, one can begin to be probabilistically explained away by the other: as the posterior probability of the model that treats the scene as a unitary object increases, the probability of the model that treats it as composite scene will go down. Once change blindness is abolished, such that both mudsplashes and scene changes are seen, a third ("Photoshop") model will have evolved on which individual components can change but not necessarily in a scene-congruent manner. All this predicts that there should be less change blindness for mudsplashes on dynamic stimuli such as movies because the causal model for such stimuli has higher accuracy; it also predicts less blindness if the mudsplashes are meaningful in the original scene such that competition between models is not engendered.

For some scene changes it is harder to induce change blindness. Mudsplashes can blind us when a rock in the way of a kayak changes into a log, but blinds us less when the rock changes into another kayak (Sampanes et al., 2008). This type of situation is often dealt with in terms of gist changes but it is also consistent with the interpretation given above. The difference between a log and another kayak in the way of the kayak is in the change in parameters of the model explaining away the prediction error. The change from an unmoving object (rock) to another unmoving object (log) incurs much less model revision than the change to a moving, intentional object (other kayak): the scope for causal interaction between two kayaks is much bigger than for one kayak and a log. The prediction error is thus much bigger for the latter, and updating the model to reflect this will increase its probability more, and make blindness less likely.

A different type of change blindness occurs when there is no distractor but the change is very slow and incremental (e.g., Simons et al., 2000), such as a painting where one part changes color over a relatively long period of time. Without attention directed at the changing property, the change is not noted. In this case it seems likely that each incremental change is within the expected variability for the model of the entire scene. When attention is directed at the slowly changing component of the scene, the precision expectation and thus the weighting goes up, and it is more likely that the incremental change will generate a prediction error. This is then an example of change blindness due to imprecise prediction error minimization. If this is right, a prediction is that change of precision expectation through learning, or individual differences in such expectations, should affect this kind of change blindness.

SHORT TERM COVERT ATTENTION ENHANCES CONSCIOUS PERCEPTION

If a peripheral cue attracts covert attention to a grating away from fixation, then conscious experience of its contrast is enhanced (Carrasco et al., 2004). Similar effects are found for spatial frequency and gap size (Gobell and Carrasco, 2005). In terms of precision, the peripheral cue induces a high precision expectation for the cued region, which increases the weighting for prediction error from the low contrast grating placed there. Specifically, the expectation will be for a stimulus with an improved signal to noise ratio, that is, a stronger signal. This then seems to be a kind of self-fulfilling prophecy: an expectation for a strong bottom-up signal causing a stronger error signal. The result is that the world is being represented as having a stronger, more precise signal than it really has, and this is then reflected in conscious perception.

From this perspective, the attentional effect is parasitic on a causal regularity in the world. Normally, when attention is attracted to a region there will indeed be a high signal to noise event in that region. This is part of the prediction error minimization role for attention described above. If this regularity did not hold, then exogenous attention would be costly in free energy. In this way the effect from Carrasco's lab is a kind of attentional visual illusion. A further study provides evidence for just this notion of an invariant relation between cue strength and expectation for subsequent signal strength: the effect is weakened as the cue contrast decreases (Fuller et al., 2009). The cue sets up an expectation for high signal strength (i.e., high precision) in the region and so it makes sense that the cue strength and the expectation are tied together. It is thus an illusion because a causal regularity about precision is applied to a case where it does not in fact hold. If it is correct that this effect relies on learned causal regularities, then it can be predicted that the effect should be reversible through learning, such that strong cues come to be associated with expectations for imprecise target stimuli and vice versa[6].

At the limit, this paradigm provides an example of attention directed at subthreshold stimuli, and thereby enabling their selection into conscious perception (e.g., 3.5% contrast subthreshold grating is perceived as a 6% contrast threshold grating (Carrasco et al., 2004). This shows nicely the modulation by precision weighting of the overall free energy landscape: prediction error, which initially is so imprecise that it is indistinguishable from expected noise can be up-weighted through precision expectations such that the internal model is eventually revised to represent it. Paradoxically, however, here what we have deemed an attentional illusion of stimulus precision facilitates veridical perception of stimulus occurrence.

[6]It is a tricky question whether or not this attentional effect is then explained without appealing to "mental paint" (Block, 2010), and whether it is therefore a challenge to representationalism about conscious perception. Precision optimization is an integral part of perceptual inference, which is all about representing the causal structure of the world. As such the explanation is representational. But it concerns precision, which is an often neglected aspect of representation: the representationalism assumed here allows that a relatively accurate representation can fail to optimize precision. What attention itself affords is improved precision, not accuracy (see Prinzmetal et al., 1997).

It is an interesting question if the self-fulfilling prophecy suggested to be in play here is always present under attention, such that attention perpetually enhances phenomenology. In predictive coding terms, the answer is probably "no." The paradigm is unusual in the sense that it is a case of *covert* attention, which stifles normal active inference in the form of fixation shifts. If central fixation is abolished and the low contrast grating is fixated, the bound on free energy is again minimized, and this time the error between the model and the actual input from the grating is likely to override the expectation for a strong signal.

This attentional illusion works for exogenous cueing but also for endogenous cueing (Liu et al., 2009), where covert endogenous attention is first directed at a peripheral letter cue, is sustained there, and then enhances the contrast of the subsequent target grating at that location. There does not seem to be any studies of the effect of endogenous attention that is entirely volitional and not accompanied by high contrast cues in the target region (even Ling and Carrasco, 2006 has high contrast static indicators at the target locations).

From the point of view of predictive coding, the prediction is then that there will be less enhancing effect of such pure endogenous attention since the high precision expectation (increased baseline) in this case is not induced via a learned causal regularity linking strong signal cues to strong signal targets.

A more general prediction follows from the idea that attention is driven by the (hyper-) prior that cues with high signal strength have high signal to noise ratio. It may be possible to revert this prior through learning such that attention eventually is attracted by low strength cues and stronger cues are ignored. In support of this prediction, there is evidence that some hyperpriors can be altered, such as the light from above prior (Morgenstern et al., 2011).

This attentional effect is then explained by precision optimization leading to an illusory perceptual inference. It is a case of misrepresented high precision combined with relatively low accuracy.

SUSTAINED COVERT ATTENTION DIMINISHES CONSCIOUS PERCEPTION AND ENHANCES FILLING-IN

In Troxler fading (Troxler, 1804) peripheral targets fade out of conscious perception during sustained central fixation. If attention but not fixation is endogenously directed at one type of sensory attribute, such as the color of some of the peripheral stimuli, then those stimuli fade faster than the unattended stimuli (Lou, 1999).

It is interesting that here attention seems to diminish conscious perception whereas in the cases discussed in the previous section it enhances it. A key factor here is the duration of trials: fading occurs after several seconds and enhancement is seen in trials lasting only 1–2 s. This temporal signature is consistent with predictive coding insofar as when the prediction error from a stimulus is comprehensively suppressed and no further exploration is happening (since active inference is subdued due to central fixation during covert attention) probability should begin to drop. This follows from the idea that what drives conscious perception is the actual process of suppressing prediction error. It translates to the notion that the system expects that the world cannot be unchanging for very long periods of time (Hohwy et al., 2008).

In Troxler fading there is an element of filling-in as the fading peripheral stimuli are substituted by the usually gray background. This filling-in aspect is seen more dramatically if the background is dynamic (De Weerd et al., 2006): as sustained attention diminishes perception of the peripheral target stimuli, it also amplifies conscious perception by illusory filling-in. A similar effect is seen in motion induced blindness (MIB). Here peripheral targets fade when there is also a stimulus of coherently moving dots, and the fading of the peripheral dots happens faster when they are covertly attended (Geng et al., 2007; Schölvinck and Rees, 2009).

The question is then why attention conceived as precision weighting should facilitate the fading of target stimuli together with enhancing filling-in in these cases. In Troxler fading with filling-in of dynamic background as well as in MIB there is an element of model competition. In MIB, there is competition between a model representing the coherently moving dots as a solid rotating disk, which if real would occlude the stationary target dots, and a model representing isolated moving dots, which would not occlude the target dots. The first model wins due to the coherence of the motion. An alternative explanation is that there is competition between a model on which there is an error (a "perceptual scotoma") in the visual system, and a model where there is not; in the former case, it would make sense for the system to fill-in (New and Scholl, 2008). In the Troxler case with a dynamic background, there is competition between models representing the world as having vs. not having gaps at the periphery, with the latter tending to win. Sustained attention increases the precision weighting for *all* prediction error from the attended region, that is, for both the target stimuli and the context in which they are shown (i.e., the dynamic background or, as in MIB, the coherently moving foreground). This context is processed not only at that region but also globally in the stimulus array and this would boost the confidence that it fills the locations of the target stimuli. This means that as the prediction error for the peripheral target stimuli is explained away, the probabilistic balance might tip in favor of the model that represents the array as having an unbroken background, or a solid moving foreground (or a perceptual scotoma).

It is thus possible to accommodate these quite complex effects of covert attention within the notion of attention as precision expectation. On the one hand, exogenous cues can engender high precision expectations that can facilitate target perception, and, on the other hand these expectations can facilitate filling-in of the target location. At the same time, covert attention stifles active inference and engenders a degree of inaccuracy.

EXOGENOUS ATTENTION TO INVISIBLE STIMULI

During continuous flash suppression, perceptually suppressed images of nudes can attract attention in the sense that they function as exogenous cues in a version of the Posner paradigm (Jiang et al., 2006). This shows that a key attentional mechanism works in the absence of conscious perception. When there are competing models, conscious perception is determined by the model with the highest posterior probability. It is conceivable that though the nude image is a state in a losing model it may still induce precision-related gain for a particular region. In general, in the processing of emotional stimuli, there is clear empirical evidence to suggest

that fast salient processing (that could mediate optimization of precision expectations) can be separated from slower perceptual classification (Vuilleumier et al., 2003). Evidence for this separation rests on the differences in visual pathways, in terms of processing speeds and spatial frequencies that may enable the salience of stimuli to be processed before their content. Even though a high precision expectation could thus be present for the region of the suppressed stimulus, it is possible for the overall prediction error landscape to not favor the generative model for that stimulus over the model for the abruptly flashing Mondrian pattern in the other eye. The result is that the nude image is not selected for conscious perception but that there nevertheless is an expectation of high precision for its region of the visual field, explaining the effect.

CONCLUDING REMARKS

The relation between conscious perception and attention is poorly understood. It has proven difficult to connect the two bodies of empirical findings, based as they are on separate conceptual analyses of each of these core phenomena, and fit them into one unified picture of our mental lives. In this kind of situation, it can be useful to instead begin with a unified theoretical perspective, apply it to the phenomena at hand and then explore if it is possible to reasonably interpret the bodies of evidence in the light of the theory.

This is the strategy pursued here. The idea that the brain is a precision-weighted hypothesis tester provides an attractive vision of the relationship. Because the states of the world have varying levels of noise or uncertainty, perceptual inference must be modulated by expectations about the precisions of the sensory signal (i.e., of the prediction error). Optimization of precision expectations, it turns out (Feldman and Friston, 2010), fits remarkably well the functional role often associated with attention. And the perceptual inference which, thus modulated by attention, achieves the highest posterior probability fits nicely with being what determines the contents of conscious perception.

In this perspective, attention and conscious perception are distinct but naturally connected in a way that allows for what appears to be reasonable and fruitful interpretations of some key empirical studies of them and their relationship. Crudely, perception and attention stand to each other as accuracy and precision, statistically speaking, stand to each other. We have seen that this gives rise to reasonably coherent interpretations of specific types of experimental paradigms. Further mathematical modeling and empirical evidence is needed to fully bring out this conjecture, and a number of the interpretations were shown to lead to testable predictions.

To end, I briefly suggest this unifying approach also sits reasonably well with some very general approaches to attention and perception.

From a commonsense perspective, endogenous and exogenous attention have different functional roles. Endogenous attention can only be directed at contents that are already conscious (how can I direct attention to something I am not conscious of?) and when states of affairs grab exogenous attention they thereby become conscious (if I fail to become aware of something then how could my attention have been grabbed?). This is an oversimplification, as can be seen from the studies reviewed above. The mapping of conscious perception and attention onto the elements of predictive coding can explain the commonsense understanding of their relationship but also why it breaks down. Normally endogenous attention is directed at things we already perceive so that no change is missed, i.e., more precision is expected and the gain is turned up. But precision gain itself is neutral on the actual state of affairs, it just makes the system more sensitive to prediction error, so if we direct attention at a location that seems empty but that has a subthreshold stimulus we are still more likely to spot it in the end. Conversely, even if precision expectations are driven up by an increase in signal strength somewhere, and attention in this sense is grabbed, it does not follow that this signal must drive conscious perception. A competing model may as a matter of fact have higher probability.

It is sometimes said that a good way to conceive of conscious perception and attention is in terms of the former as a *synthesizer* that allows us to make sense of our otherwise chaotic sensory input, and the latter as an *analyzer* that allows us to descend from the overall synthesized picture and focus on a few more salient things (Van Boxtel et al., 2010). The predictive coding account allows this sentiment: prediction error minimization is indeed a way of solving the inverse problem of figuring out what in the world caused the sensory input, and attention does allow us to weight the least uncertain parts of this signal. The key insight from this perspective is however that though these are distinct neural processes they are both needed to allow the brain to solve its inverse problem. But when there are competing models, they can work against each other, and conscious perception can shift between models as precisions and bounds are optimized and the world selectively sampled.

Perhaps the most famous thing said about attention is from James:

> Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatter-brained state which in French is called *distraction*, and *Zerstreutheit* in German (James, 1890, Vol. I, pp. 403–404).

The current proposal is that "attention is simply the process of optimizing precision during hierarchical inference" (Friston, 2009, p. 7). This does not mean the predictive coding account of attention stands in direct opposition to the Jamesian description. It is a more accurate, reductive and unifying account of the mechanism underlying parts of the phenomenon James is trying to capture: James' description captures many of the aspects of endogenous attention and model competition that are discussed in terms of precision in this paper.

The sentiment that attention is intimately connected with perception in a hypothesis testing framework was captured very early on by Helmholtz. He argued, for example, that binocular rivalry is an attentional effect but he explicated attention in terms of

activity, novelty, and surprise, which is highly reminiscent of the contemporary predictive coding framework:

> The natural unforced state of our attention is to wander around to ever new things, so that when the interest of an object is exhausted, when we cannot perceive anything new, then attention against our will goes to something else. [. . .] If we want attention to stick to an object we have to keep finding something new in it, especially if other strong sensations seek to decouple it (Helmholtz, 1860, p. 770; translated by JH).

Helmholtz does not here mention precision expectations but they find a natural place in his description of attention's role in determining conscious content: precision expectations enable attention to stick, where sticking helps, and to wander more fruitfully too.

ACKNOWLEDGMENTS

Thanks to Karl Friston, Tim Bayne, and the reviewers for very helpful comments and suggestions. This research is supported by the Australian Research Council.

REFERENCES

Badcock, P. B. (2012). Evolutionary systems theory: a unifying meta-theory of psychological science. *Rev. Gen. Psychol.* 16, 10–23.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci. (Regul. Ed.)* 11, 280–289.

Bayne, T. (2010). *The Unity of Consciousness.* Oxford: Oxford University Press.

Bayne, T., and Montague, M. (2011). *Cognitive Phenomenology*. Oxford: Oxford University Press.

Bays, P. M., and Wolpert, D. M. (2007). Computational principles of sensorimotor control that minimize uncertainty and variability. *J. Physiol. (Lond.)* 578, 387–396.

Block, N. (1995). On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–287.

Block, N. (2008). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* 30, 481–499.

Block, N. (2010). Attention and mental paint. *Philos. Issues* 20, 23–63.

Boly, M., Garrido, M. I., Gosseries, O., Bruno, M.-A., Boveroux, P., Schnakers, C., Massimini, M., Litvak, V., Laureys, S., and Friston, K. (2011). Preserved feedforward but impaired top-down processes in the vegetative state. *Science* 332, 858–862.

Brown, H., Friston, K. J., and Bestmann, S. (2011). Active inference, attention and motor preparation. *Front. Psychol.* 2:218. doi:10.3389/fpsyg.2011.00218

Carrasco, M., Ling, S., and Read, S. (2004). Attention alters appearance. *Nat. Neurosci.* 7, 308–313.

Casella, G. (1992). Illustrating empirical Bayes methods. *Chemometr. Intell. Lab. Syst.* 16, 107–125.

Chabris, C. F., Weinberger, A., Fontaine, M., and Simons, D. J. (2011). You do not talk about fight club if you do not notice fight club: inattentional blindness for a simulated real-world assault. *i-Perception* 2, 150–153.

Chalmers, D. (1996). *The Conscious Mind.* Harvard: Oxford University Press.

Cohen, M. A., Alvarez, G. A., and Nakayama, K. (2011). Natural-scene perception requires attention. *Psychol. Sci.* 22, 1165–1172.

Cohen, M. A., and Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends Cogn. Sci. (Regul. Ed.)* 15, 358–364.

De Weerd, P., Smith, E., and Greenberg, P. (2006). Effects of selective attention on perceptual filling-in. *J. Cogn. Neurosci.* 18, 335–347.

Dennett, D. C. (1991). *Consciousness Explained.* Boston: Little, Brown & Co.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1245.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193.

Ditchburn, R. W., and Ginsborg, B. L. (1952). Vision with a stabilized retinal image. *Nature* 170, 36–37.

Feldman, H., and Friston, K. (2010). Attention, uncertainty and free-energy. *Front. Hum. Neurosci.* 4:215. doi:10.3389/fnhum.2010.00215

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol.* 4, e1000211. doi:10.1371/journal.pcbi.1000211

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci. (Regul. Ed.)* 13, 293–301.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138.

Friston, K., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biol. Cybern.* 104, 137–160.

Friston, K., and Stephan, K. (2007). Free energy and the brain. *Synthese* 159, 417–458.

Friston, K. J., Daunizeau, J., and Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS ONE* 4, e6421. doi:10.1371/journal.pone.0006421

Friston, K. J., Thornton, C., and Clark, A. (in press). Free-energy minimization and the dark room problem. *Front. Psychol.*

Fuller, S., Park, Y., and Carrasco, M. (2009). Cue contrast modulates the effects of exogenous attention on appearance. *Vision Res.* 49, 1825–1837.

Geng, H., Song, Q., Li, Y., Xu, S., and Zhu, Y. (2007). Attentional modulation of motion-induced blindness. *Chin. Sci. Bull.* 52, 1063–1070.

Gobell, J., and Carrasco, M. (2005). Attention alters the appearance of spatial frequency and gap size. *Psychol. Sci.* 16, 644–651.

Goodman, N. (1955). *Fact, Fiction and Forecast.* Cambridge, MA: Harvard University Press.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 181–197.

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci. (Regul. Ed.)* 10, 14–23.

Harrison, L., Bestmann, S., Rosa, M. J., Penny, W., and Green, G. G. R. (2011). Time scales of representation in the human brain: weighing past information to predict future events. *Front. Hum. Neurosci.* 5:37. doi:10.3389/fnhum.2011.00037

Helmholtz, H. V. (1860). *Treatise on Physiological Optics.* New York: Dover.

Hesselmann, G., Sadaghiani, S., Friston, K. J., and Kleinschmidt, A. (2010). Predictive coding or evidence accumulation? False inference and neuronal fluctuations. *PLoS ONE* 5, e9926. doi:10.1371/journal.pone.0009926

Hillyard, S. A., and Mangun, G. R. (1987). Sensory gating as a physiological mechanism for visual selective attention. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 40, 61–67.

Hohwy, J. (2009). The neural correlates of consciousness: new experimental approaches needed? *Conscious. Cogn.* 18, 428–438.

Hohwy, J., and Fox, E. (2012). Preserved aspects of consciousness in disorders of consciousness: a review and conceptual analysis. *J. Conscious. Stud.* 19, 87–120.

Hohwy, J., Roepstorff, A., and Friston, K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition* 108, 687–701.

Hume, D. (1739–1740). *A Treatise of Human Nature.* Oxford: Oxford: Clarendon Press.

Jack, C. E., and Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Percept. Mot. Skills* 37, 967–979.

James, W. (1890). *The Principles of Psychology*. New York: Holt.

Jiang, Y., Costello, P., Fang, F., Huang, M., and He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proc. Natl. Acad. Sci. U.S.A.*103, 17048–17052.

Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55, 271–304.

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of timescales and the brain. *PLoS Comput. Biol.* 4, e1000209. doi:10.1371/journal.pcbi.1000209

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2010). Perception and hierarchical dynamics. *Front. Neuroinformatics* 4, 12.

Kitcher, P. (1989). "Explanatory unification and the causal structure of the world," in *Scientific Explanation*, eds P. Kitcher and W. Salmon (Minneapolis: University of Minnesota Press), 410–505.

Koch, C., and Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends Cogn. Sci. (Regul. Ed.)* 11, 16–22.

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., and De Lange, F. P. (2011). Attention reverses the effect of prediction in silencing sensory signals. *Cereb. Cortex*. doi: 10.1093/cercor/bhr310

Kveraga, K., Ghuman, A. S., and Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain Cogn.* 65, 145–168.

Ling, S., and Carrasco, M. (2006). When sustained attention impairs perception. *Nat. Neurosci.* 9, 1243–1245.

Liu, T., Abrams, J., and Carrasco, M. (2009). Voluntary attention enhances contrast appearance. *Psychol. Sci.* 20, 354–362.

Lou, L. (1999). Selective peripheral fading: evidence for inhibitory sensory effect of attention. *Perception* 28, 519–526.

Martinez-Trujillo, J. C., and Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.* 14, 744–751.

McGrayne, S. B. (2011). *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. New Haven: Yale University Press.

Morgenstern, Y., Murray, R. F., and Harris, L. R. (2011). The human visual system's assumption that light comes from above is weak. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12551–12553.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol. Cybern.* 66, 241–251.

New, J. J., and Scholl, B. J. (2008). Perceptual scotomas. *Psychol. Sci.* 19, 653–659.

Prinzmetal, W., Nwachuku, I., Bodanski, L., Blumenfeld, L., and Shimizu, N. (1997). The Phenomenology of Attention. *Conscious. Cogn.* 6, 372–412.

Rao, R. P. N. (2005). Bayesian inference and attentional modulation in the visual cortex. *Neuroreport* 16, 1843–1848.

Rensink, R., O'Regan, J., and Clark, J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.* 8, 368.

Ross, J., and Burr, D. (2008). The knowing visual self. *Trends Cogn. Sci. (Regul. Ed.)* 12, 363–364.

Sampanes, A. C., Tseng, P., and Bridgeman, B. (2008). The role of gist in scene recognition. *Vision Res.* 48, 2275–2283.

Schölvinck, M. L., and Rees, G. (2009). Attentional influences on the dynamics of motion-induced blindness. *J. Vis.* 9. (Article 38):1–8.

Simons, D. J., and Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* 28, 1059–1074.

Simons, D. J., Franconeri, S. L., and Reimer, R. L. (2000). Change blindness in the absence of a visual disruption. *Perception* 29, 1143–1154.

Summerfield, C., and Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends Cogn. Sci. (Regul. Ed.)* 13, 403–409.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., and Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* 11, 1004–1006.

Todorovic, A., Van Ede, F., Maris, E., and De Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J. Neurosci.* 31, 9118–9123.

Treisman, A. (1996). The binding problem. *Curr. Opin. Neurobiol.* 6, 171–178.

Treisman, A. (1998). Feature binding, attention and object perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1295–1306.

Treisman, A. (1999). Solutions to the binding problem: review progress through controversy summary and convergence. *Neuron* 24, 105–110.

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136.

Troxler, D. (1804). "Über das Verschwindern gegebener Gegenstande innerhalb unsers Gesichtskreises," in *Ophthalmologisches Bibliothek*, eds K. Himly and J. A. Schmidt (Jena: Fromman), 1–119.

Van Boxtel, J. J. A., Tsuchiya, N., and Koch, C. (2010). Consciousness and attention: on sufficiency and necessity. *Front. Psychol.* 1:217. doi:10.3389/fpsyg.2010.00217

Vuilleumier, P., Armony, J. L., Driver, J., and Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat. Neurosci.* 6, 624–631.

Watzl, S. (2011a). The nature of attention. *Philos. Compass* 6, 842–853.

Watzl, S. (2011b). The philosophical significance of attention. *Philos. Compass* 6, 722–733.

Wu, W. (2011). Confronting many-many problems: attention and agentive control. *Noûs* 45, 50–76.

Zhang, P., Jamison, K., Engel, S., He, B., and He, S. (2011). Binocular rivalry requires visual attention. *Neuron* 71, 362–369.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 December 2011; paper pending published: 18 January 2012; accepted: 14 March 2012; published online: 02 April 2012.

*Citation: Hohwy J (2012) Attention and conscious perception in the hypothesis testing brain. Front. Psychology **3**:96. doi: 10.3389/fpsyg.2012.00096*

This article was submitted to Frontiers in Consciousness Research, a specialty of Frontiers in Psychology.