7

The Many Faces of Attention: Why Precision Optimization Is Not Attention

Madeleine Ransom
Indiana University Bloomington, Department of Cognitive Science
Sina Fazelpour
Carnegie Mellon University, Department of Philosophy

1. Introduction

Predictive coding in its most general form is the view that a process reduces the amount of information that needs to be stored or transmitted by utilizing predictions. This reduction of information is often understood in terms of prediction error; the process represents only the difference between the predicted and actual input instead of representing the input directly.[1] It thus presupposes that the process employs a generative model: there must be a way to generate predictions so that they can be compared to incoming input, as well as a way to update these predictions in light of prediction error.

Predictive coding as it has been developed by Karl Friston (2008, 2009, 2010), and adopted by philosophers such as Andy Clark (2013a, 2015) and Jakob Hohwy (2012, 2013), is claimed to be a grand unifying theory of all cognitive functioning, though it has been most thoroughly developed as a theory of perception. Predictive coding combines the idea of transmitting only prediction error and using a generative model to make predictions with two other key concepts: Bayesian inference and a hierarchically structured model.[2] From here on, we shall refer to this combination of concepts as predictive coding (PC), and it is PC that our discussion and criticism here applies to.

The generative model as it is conceived of by PC is a model of the world—objects, forces and their interactions, people and their mental states. Possessing an internal generative model is key to addressing the problem our perceptual systems face: they must infer the distal causes (things out in the world) of sensory input, but there is not a one to one mapping between these two. The same sensory input could be caused by distinct objects, and a single object could produce different sensory inputs. A model generates hypotheses as to the potential distal causes of sensory input, and it also constrains the hypothesis space by limiting the sorts of hypotheses that can be generated.

Bayesian inference provides guidance on how to update hypotheses in light of new information. On Bayesian inference, the totality of the hypotheses produced by the generative model is the hypothesis space, with each hypothesis assigned a subjective prior probability. We also generate hypotheses of the likelihood of a given sensory input given certain worldly conditions. Upon receipt of sensory input, Bayesian conditionalization is used to update prior probabilities assigned to each hypothesis.

Due to computational intractability, the brain cannot implement a Bayesian model of the complexity needed to explain perceptual and mental processes. In light of this, PC holds that the brain instead approximates such inference by giving the generative model a hierarchical structure. The model is envisioned as being composed of distinct levels, each concerned with generating and updating predictions about the world. Higher-level hypotheses will tend to be more global predictions, which will then generate and constrain predictions at lower levels of the hierarchy. For example, a high-level prediction might be that there is a cat on a mat, and lower levels will concern the texture of the cat's fur, the color of the mat, and so on (Hohwy 2013: 27).

Putting all of these concepts together, the emerging picture (for perceptual inference) is as follows: We possess an internal model of the world that allows us to generate high-level hypotheses about the distal causes of sensory data, which in turn generate lower-level sub-hypotheses in a descending hierarchical fashion, with the more fine-grained perceptual details relegated to the bottom levels. These hypotheses are then compared to the incoming sensory data. When there is a mismatch, a prediction error is produced. These errors are the only feedforward or "bottom-up" signals propagated through the hierarchy—unresolved prediction error gets sent up to the next level, where the higher-level hypothesis is revised in a Bayesian fashion until prediction error is successfully minimized.[3] The hypotheses with the least prediction error are those selected to represent the world as we see it. This iterative process of hypothesis generation and prediction error minimization is taken to be a parsimonious and unified account of the functioning of all cognitive phenomena, including thought. Prediction error minimization is "all the brain ever does" (Hohwy 2013: 7).

A problem for PC as laid out thus far is that it must discern between noise and prediction error. When a model generates a predictive hypothesis, it isn't expected that it will fit the data exactly—there will always be some data points that don't behave according to the model's predictions. This is a problem because there are two ways data can fail to fit the hypothesis. The first is that the data is noise. In this case, the model should not be revised to accommodate these extra data points because this would lead to less accurate predictions and therefore less accurate representation of the external world. The second is that such data are legitimate prediction error. In this case, the model should be revised in light of this extra data because this will lead to more accurate predictions and so more accurate perceptual representations. The problem of uncertainty is how to discern between the two cases—when should the model be revised in light of the data, and when should it be ignored?

In order to address this problem, PC introduces second-order statistics into the model. While the measurement of prediction errors involves comparing the means of two probability distributions, expected precision is a measurement of the variance of the data about the mean. This is a measure of how reliable, or precise, we take the prediction error signal to be in a given context: how likely is it in a given situation that the incongruent data constitutes legitimate prediction error as opposed to noise? Noise in the world is context-dependent—in different contexts there will be more or less noise woven into signal. Listening to your friend's story in a deserted café is quite different from the experience of listening to her at a noisy party. However, one can begin to predict how much noise to expect in a given context on the basis of learned statistical regularities. In this way, we form precision expectations. We come to expect prediction error to be more or less precise in a given context, and so either trust it in revising the relevant hypothesis or discount it as noise and so refrain from revising the hypothesis.

Synaptic gain is thought to be the neural mechanism that encodes precision in the brain (Friston 2008). On PC, when prediction error is expected to be reliable this results in the gain on that prediction error being amplified, and so passed up the hierarchy to the level above it, where it is used to revise the hypothesis until it can successfully minimize the error, usually an iterative process. When a prediction error is expected to be unreliable, this results in the opposite effect—the prediction error is suppressed, or dampened. In this case the prediction error does not initiate revision of the hypothesis that generated it. Given that the total gain on prediction error must always sum to one, expected precisions drive selective processing by weighting some prediction errors over others (Hohwy 2012: 8). Such selectivity is often thought to be the functional role of attention, and so provides the basis for the PC theory of attention.

The PC theory of attention proposed by Friston (2009) and Feldman and Friston 2010) and defended by Hohwy (2012, 2013) and Clark (2013a, 2015) is that "attention is simply the process of optimizing precision during hierarchical inference" (Friston 2009: 299).

Given the goal of minimizing prediction error, the optimal way in which to revise perceptual hypotheses is on the basis of the prediction errors with the highest expected precisions. Expending effort on prediction errors likely containing high noise levels will lead to suboptimal perceptual results. Better to focus on the signals that are more likely to lead to improvements in the model—the reliable signals.

Attention then falls out directly from the PC account of perception. There is no machinery involved in the process over and above that which is needed to account for perceptual inference. On this picture attention has the functional role of guiding perceptual inference by directing processing resources toward the prediction errors with the higher expected precisions given a context. Again, this results in the minimization of prediction error, though attention is concerned only with expected precision of prediction error and not directly with the accuracy of hypotheses. However, because the estimation of expected precisions is a fundamental aspect of perceptual inference, then so is attention.

While there are overarching theories of attention that provide a unitary explanation of the phenomenon in terms of cognitive unison (Mole 2011), selection for action (Wu 2014), or some type of prioritizing activity (Watzl 2017), there are nevertheless important differences between different varieties of attention. There is a phenomenological difference between having your attention grabbed by a loud explosion on the street, and deciding to attend to the words on this page. Empirical researchers have also identified substantive behavioral and neural differences between different kinds of attention, such as time course and involved brain regions (Carrasco 2011, Posner 2011). In order for the PC claim that attention is the optimization of precision expectations to go through, then it must be able to accommodate all varieties of attentional phenomena. Here we review those varieties that have been given the most thorough treatment in the PC literature. The primary focus of PC has been on explaining both sides of a canonical division of attention into endogenous and exogenous processes (Posner 1980).

Exogenous attention is an automatic orienting response to environmental stimuli. Intuitively, it is when attention is captured by some object or state of affairs that the agent was not already previously searching for. It can thus be described in terms of task relevance: while the object in question is irrelevant to the task the agent is currently performing, it nevertheless captures her attention. This is thought to occur sometimes due to contextually mediated perceptual salience—whether something is salient or not will depend upon the context it is presented in. A neon pink circle against a white backdrop catches the eye; against a neon pink backdrop it does not.

The PC account of exogenous attention is as follows. First, one's generative model must formulate a hypothesis about a region of space or an object that one is not currently attending to (or that is not the main focus of attention). Note that because attention sets the gain on prediction error, there will therefore be a low gain on the prediction error generated for unattended or peripherally attended regions or objects. Second, the presentation of a stimulus such as a sudden noise, light, or movement results in an abrupt and large prediction error. Third, the gain on this prediction error will be amplified. This is because of a learned statistical regularity (precision expectation) that in our sort of environment any change in variability due to noise is rather slow and small, so the abrupt onset of the signal itself is taken as an indication of its high precision (Hobson and Friston 2012: 92, Hohwy 2013: 197). Given this precision expectation, the gain or amplitude of the already large prediction error will be enhanced; this amounts to paying attention to the stimulus. Fourth, attention will then cause the hypothesis to be revised preferentially in light of this prediction error.

In endogenous attention, the object or spatial region of attention is to some degree directed by the agent, usually according to a purpose or task. Empirical studies of attention, including those conducted within PC, tend to operationalize endogenous attention as task relevance (Chennu et al. 2013, Jiang et al. 2013, Kok et al. 2011). PC explains endogenous attention by positing that we have learned precision expectations that guide such endogenous shifts. Consider the classic Posner paradigm (Posner 1980), which is a cueing task that investigates how covert attention can facilitate stimulus detection. In one common version of the task an arrow provides a spatial cue as to where the stimulus will likely appear. Through repeated trials the subject learns that when an arrow is shown pointing to a given area on a computer screen, an object will likely appear in that area. This learned statistical regularity is a contextually mediated precision expectation: when there is an arrow pointing toward a given location, the prediction error that will subsequently be produced by the appearance of the object in that location is expected to be precise, or reliable. When an arrow appears on the screen, pointing to the bottom right corner, this causes the gain from the prediction error issuing from this region to be increased (this is tantamount to saying that one pays attention to the bottom right corner). As a consequence, when the stimulus appears it is perceived more rapidly.

With these explanations of exogenous and endogenous attention on the table, we can begin to build our criticism. In section 2 we argue that the weighting of prediction error based on expected precisions is too narrow a phenomenon to be identified with attention, because it cannot accommodate the full range of attentional phenomena. We review criticisms that PC cannot account for volitional attention and affect-biased attention, and we propose that it also cannot account for intellectual attention and feature-based attention. In section 3 we argue that the precision weighting of prediction error is too broad a phenomenon to be identified with attention because such weighting plays a central role in multimodal integration.

2. PC Theory of Attention Is too Narrow

The varieties of attentional phenomena are richer than a simple bipartite division into endogenous and exogenous attention. Indeed, the very usefulness of this dichotomy in the face of empirical research has come under question (Awh et al. 2012, Todd and Manaligod 2018). Even operating with the division in place, however, further distinctions can be made between spatial, object, and feature-based attention, as well as overt and covert attention, among others (Carrasco 2011, Mole 2011, Wu 2014). In what follows, we don't cover the full range of divisions that have been made, but instead focus on those that have been flagged as problematic for PC—affect-biased attention and volitional attention—and then propose that intellectual attention and feature-based attention are also problematic.

2.1 Affect-biased Attention

Affect-biased attention is attention to stimuli that are affectively salient in virtue of their association with reward or punishment (Markovic et al. 2014, Mather and Sutherland 2011, Rolls 2000, Todd and Manaligod 2018, Vuilleumier 2015). As Ransom et al. (2020) point out, it is not straightforwardly categorized as exogenous attention because affectively salient objects can succeed in capturing attention in spite of a lack of physical salience (Anderson et al. 2011, 2012, Anderson 2013, Anderson and Yantis 2013, Della Libera and Chelazzi 2009, Hickey Chelazzi and Theeuwes 2010, Niu et al. 2012, Shomstein and Johnson 2013). Neither does it sit comfortably in the category of endogenous attention, given that affectively salient objects also capture attention when they are not task relevant (Awh et al. 2012, Todd et al. 2012).

The principle criticism that Ransom et al. (2020) raise, however, is that precision expectations and affective salience are dissociable. There will be occasions upon which we will have low-precision expectations for prediction errors generated by a given hypothesis, but where the potential rewards or punishment is significant enough that we ought nevertheless to attend. For example, if you are hiking in the Pacific

Northwest, a rustle in the bush is most likely caused by wind or a bird, but there is a small chance it could be a large predator such as a cougar or bobcat. In such cases, the cost of getting it wrong is so high that one ought to attend, regardless of the fact that prediction error generated by the movement in the grass is likely best dismissed as noise. Examples such as these suggest that it is not just the expected precision of the prediction error, but also the value or affective salience of the object that ought to drive attention. Affective salience thus constitutes an independent source of influence, one that is not assimilable to precision expectations (see also Colombo and Wright 2017, Gershman and Daw 2012).

2.2. Volitional Attention

While endogenous attention is driven by learned statistical regularities pertaining to the external environment, sometimes we attend to things out of will or whimsy. If I make up my mind to pay attention to the tree outside my window, or the bare wall of my office, then I am generally able to do so. This ability isn't well captured by the above account of endogenous attention, where the top-down cueing is still relatively automatic, and is based on an external prompt. One pays attention to the spatial region indicated by the arrow without deciding to do so in any substantive, volitional sense. In order to account for volitional attention, Hohwy invokes active inference (2013: 197–9).

Active inference is meant to explain how and why it is that we perform intentional actions, using the explanatory tools of PC. Rather than mere passive perceivers, we move through our environment and act so as to change it. On PC, this is the result of generating a counterfactual hypothesis (where the content is the desired state of the organism), and then acting in the world to bring about the desired state and so minimize prediction error with respect to the hypothesis. While both perceptual and active inference have the goal of minimizing predictive error, in perceptual inference the hypothesis is revised on the basis of the prediction error, and in active inference the hypothesis is held fixed and the agent's position in the world is altered such that the resulting incoming sensory data aligns with the hypothesis. This prediction error will be in part proprioceptive—it will pertain to one's sense of one's body position in space.

For example, suppose you are currently sitting in your armchair and, given your desire to go look out the window, you generate the counterfactual hypothesis that you are standing at the window. This generates a large prediction error—you are in fact presently sitting, not standing. The way in which you will minimize prediction error is then by plotting a course of action, again using a generative hierarchical model to formulate hypotheses about the best route to take and to monitor the movement process such that prediction error with respect to course is also minimized. So you implicitly plan to take the direct path to the window by putting one foot in front of the other, and should you run into an unexpected obstacle, such as a computer extension cord, you can avoid it (minimize prediction error with respect to your plan of action) by lifting your foot higher than normal to step over it. Finally, after countless Bayesian calculations, you arrive at your goal of looking out the window. This results in a minimization of the prediction error with respect to the counterfactual hypothesis—you are now in the location you desired to be in.

Active inference is well suited to explain volitional endogenous attention because this involves a decision or desire on the part of the agent to act, though it is "a slightly unusual instance of action because the way we change our relation to the world is to increase the sensory gain in one region of space" (Hohwy 2013: 198). However, if a unitary explanation of attention is to be offered, and active inference constitutes some cases of attention, then it too must be explicable in terms of expected precisions.

The PC account of volitional attention is as follows: when we make the decision to attend to a certain spatial region, this decision functions much like the arrow cue in the Posner case—our decision causes us to expect a high-precision stimulus to occur in the relevant spatial region. This is so in virtue of a learned statistical regularity: when we choose to attend to a given region, high-precision stimuli are often detected there. Just as in the Posner case, gain from this region is then enhanced, which is tantamount to attending to it. This speeds subsequent detection of any object at the location.

Against this proposal, Ransom et al. (2017) have argued that volitional attention is not satisfactorily accommodated by PC theory of attention. This is because this same type of regularity holds for *any* stimulus whatsoever. If I decide to attend to the wall of my office, I enhance the precision of prediction errors associated with the wall. If I decide to attend to the tiny bit of lint stuck to my sweater, I get a similar result. Indeed, it is a trivial observation, and one which is taken as a datum by any account of attention: more often than not, attending to a stimulus results in enhanced precision of the signal arriving from that stimulus. The precision enhancement of such an act is a "self-fulfilling prophecy" for just about anything we decide to look at. With this in mind, precision expectations cannot be what drive attention in such cases because we will have equivalent precision expectations for all objects regardless of whether we desire to shift our attention to them. Thus, in instances such as this the optimization of precision expectations cannot be identified with attention. What really seems to be driving our attention in this case is our antecedent desire to attend, not the precision expectation.

Clark (2017) has responded to this criticism, arguing that it can be addressed through a proper appreciation of PC's recasting of desires as beliefs. This assimilation is an endorsement of what Clark has termed elsewhere "desert landscape" theories of PC (Clark 2013a). It is required in order to make good on the claim that PC is a unified theory of all cognitive functioning. On this theory, desires are understood using the PC account of active inference, as explained briefly above. They are hypotheses that are currently false of the world, but can be made true through action (Friston et al. 2011). In this sense, all (realizable) desires are self-fulfilling prophecies on PC. On this account the sorts of desires involved in volitional attention are just more of the same, and so not problematic. While there is no alteration of precision estimations out in the world, the formulation of the desire (hypothesis) together, perhaps, with changes in interoceptive context (Clark 2017: 118) are sufficient to alter the precision expectations for that hypothesis.

However, this response does not address the point that mental actions are importantly different from physical actions in that they are not resolved by minimizing proprioceptive prediction error (see Ransom et al. 2017: 109). The problem can be made clearer by considering covert attention. Covert attention is attending to a stimulus without saccading toward it. Covert attention can be voluntary. This is easy to experience: simply keep your gaze fixated on a point in front of you, and shift only your attention to some object off to the side—a lamp or a door perhaps. It stands in contrast to overt attention, which involves shifting one's gaze and saccading to the thing or area one is interested in. Lest covert attention seem like a fringe form of attention, it has been found to precede and guide the eye movements involved in overt attention in a variety of everyday tasks—the phenomenon is ubiquitous (Carrasco 2011).

Endogenous or volitional covert attention is a pure mental adjustment of attention; it does not involve changes in the physical position of the body. This makes it unlike other cases of desire fulfillment, because there is no proprioceptive prediction error. When I desire to drink a cup of tea, a proprioceptive and perceptual prediction error is generated because I am not in fact doing so. When I desire to attend to the lamp at the periphery of my vision, keeping my eyes locked on the page in front of me, there is no corresponding proprioceptive prediction error.

It is also not clear that there are perceptual prediction errors in some cases of covert attentional shifts. If I am attending to the words on the page but desire to attend to the lamp on the corner of my desk that is still in the periphery of my vision, then the (counterfactual) perceptual hypothesis will be something like "I am looking at my lamp." But it's already true that I perceive the lamp; it's in my field of vision. So the hypothesis is more adequately described as "I am attending to the lamp." This is false, and so "desire-like" in the sense required to initiate active inference, but it's unclear both whether this hypothesis is perceptual (because we are already perceiving the lamp), and what it would mean to expect the prediction error generated by this hypothesis to be precise (because it seems to be a first-order attentional hypothesis, and so it is bizarre to think we have precision expectations for attention).

In summary, if PC analyzes volitional attention in terms of active inference, then precision expectations will pertain in part to proprioceptive prediction errors. However, in the case of covert voluntary attention there are no such prediction errors, and it is not clear whether there are perceptual prediction errors in all such cases.

Clark also responds to Ransom, Fazelpour, and Mole's criticism by suggesting that the attentional shifts involved in mental action are no more mysterious than how desires are produced in the first place.

> Whatever set of personal and environmental circumstances might conspire to install or suddenly foreground a desire (for example, the desire to attend to a [particular object]), those same circumstances are now called upon to install or suddenly foreground the behaviorally-equivalent belief.
>
> (Clark 2017: 118)

So, if other models of volitional attention have a preferred explanation that cites the relevant set of personal and environmental circumstances, the PC theorist can simply re-formulate that explanation in terms of a precision-concerning belief. And if the alternative accounts lack such an explanation, then surely we cannot view the lack of an explanation as a shortcoming of the PC account in particular.

However, claiming that there can be a post-hoc reformulation of every explanation into the formalism of PC framework is one thing. Demonstrating that a satisfactory model of attention—capable of yielding novel predictions and explanations—emerges from the PC framework using only the explanatory tools of PC is quite another thing. Given the generality of PC's formalism, the first claim is rather trivial. The challenge for the PC theorists, therefore, is to provide evidence for the second claim. This challenge is not addressed by Clark's response to Ransom, Fazelpour, and Mole's criticism.

2.3. Intellectual Attention

We have complex inner lives made up of deliberations, ruminations, mind-wanderings, memories, and imaginings, and we often attend to what goes on within. Here we will call attention to thought, broadly construed "intellectual attention" (Fortney, forthcoming). Intellectual attention can occur simultaneously with perception—one needn't close one's eyes in order to pay attention to one's thoughts. The picture is rendered more complicated insofar as sometimes these shifts are exogenous and sometimes they are endogenous—sometimes we decide to pay attention to our thoughts and sometimes our attention is drawn involuntarily inwards. A complete theory of attention should be able to account for these shifts between our inner and outer lives, as well as the shifts between different sorts of mental activities, such as when task-directed thought morphs into rumination. The PC theory of attention has had little to say so far about intellectual attention, but extrapolating from their treatment of perceptual attention, the proposal would be that we pay attention to our thoughts, imaginings, or memories when we expect them to be precise. This might take the form of a large abrupt prediction error, as in the case of exogenous attention. Perhaps the sudden thought that one has left the stove on at home is like this. Or, we may turn our attention relatively deliberately to our thoughts after an external prompt we have learned to reliably associate with our own mental activity, as with endogenous attention. A friend asks, "*What was the name of the author of Don Quixote?*" Searching for the answer to this question is accompanied by a phenomenal shift inwards—we expect to retrieve the name from memory, if anywhere.[4]

Here the challenge for the PC theory is to explain the attentional shifts in terms of precision expectations—it must provide an explanation of what sorts of regularities drive the formation of precision expectations for thoughts. This project will rest on providing an adequate theory of how thoughts—the objects of intellectual attention—are generated in the first place. This is no small task. Recall the fundamental role that sensory feedback systems play in hypothesis revision for perceptual and active inference. In the case of perception, the world serves as the hidden cause of the sensory effects that drive perceptual hypothesis revision. Perceptions are predictions of the sensory effects of the world, prediction errors tell us when our predictions are off, and precision expectations tell us whether we should alter our predictions to resolve the prediction error or not. Without a hidden cause that produces the sensory effects that provide feedback to the model, the hypothesis generation process stands uncorrected, and the model becomes "top-heavy," producing ever-wilder hypotheses without revisions. In the perceptual case, this results in visual hallucinations. What then—if anything—serves as the hidden cause in the case of the mental, constraining our thoughts and preventing them from being anything other than pure fantastical fabrications?

While some of our stranger thoughts and imaginings might be well explained by the relative lack of bottom-up prediction error, we also create other more "grounded" thoughts that stay on track, or have some relatively stable connection with the world. PC may be able to explain some of these more grounded thoughts and imaginings in terms of offline proprioceptive simulations that have been used to account for long-term planning. Clark (2013b) follows Friston et al. (2011) in suggesting that precision weighting may play a role in allowing us to run simulations of future actions without actually carrying them out. By assigning a low weighting to the prediction error from lower levels of the motor system, one can simulate carrying out some action without actually bringing it about.

The precision weighting of prediction error will also likely play a role in explaining how it is that we shift between different kinds of mental activities. Clark (2013b: 5–6) follows Daw et al. (2005) in employing precision-weighted prediction error to explain how we switch from model-based to model-free decision strategies (roughly, strategies that directly involve drawing on the resources of the generative model posited by PC to problem-solve and those that involve only learning action-value pairings directly). When we are confronted with a given problem, there may be multiple competing neural resources that are employed to formulate different action plans. Which of these plans is selected will be determined by which has the highest expected precisions.

Perhaps a similar story can be told for other varieties of thought: for example, we may come to trust the deliverances of memory over imagining in some contexts. This picture would need to explain how it is that our thoughts are sometimes derailed by other systems in a given context. For example, goal-directed thought, such as trying to understand the point an author is making in a philosophical paper, may devolve into mind-wandering over time (Christoff 2012).

At this point, intellectual attention and understanding our mental lives more generally in terms of PC constitute not so much a problem for PC but rather an unexplored frontier. Whether PC theory can accommodate intellectual attention, and the full range of mental phenomena, will depend on details that have yet to be specified. However, the role that affective salience likely plays in directing our attention inwards suggests that the same issue that was raised for perceptual attention in section 2.1 will recur here. Consider again the thought that you may have left the stove on. Supposing that you are not in general a forgetful person, then it's unlikely that this is the case. But the consequences of having left the stove on are potentially dire, and so it seems that the thought ought to capture our attention regardless of low associated precision expectations. If this is right, then intellectual attention may also be driven by affective salience independently of precision expectations.

2.4. Feature-based Attention

Within the category of endogenous attention, several theorists make a further distinction between spatial attention and feature-based attention (Carrasco 2011, Egner et al. 2008, Rossi and Paradiso 1995). We can attend to a given spatial region due to some learned regularity such as that an arrow cue predicts that something will appear in a given location. Or, we can attend to a given feature of an object because we have learned that it provides relevant information, such as when we attend to the eyes and mouth areas of a person in order to assess their emotional state. Feature-based attention guides visual search, such as when while waiting for a friend we look for particular features that may distinguish her from the crowd (e.g., her height, the color of her jacket, and so on).

In the Posner paradigmdescribed above, the arrow cue does not predict which object with what type of features will appear (unless a hypothesis has been formed for this as well via conditioning, such as that dots are likely to appear after arrow cues). However, Hohwy claims that the same sort of explanation given for spatial endogenous attention can be applied to feature-based endogenous attention (2013: 196). In such cases, we can focus our attention on particular features of an object in a scene. Supposing we are looking to identify all the red apples in a barrel, or all the people with red and white striped shirts in a crowd. Feature-based endogenous attention allows us to do so quickly and efficiently, even if the spatial location of these features is not well predicted. How might this work? It is crucial that it do so, as many cases of endogenous attention are those involving searches for certain features or objects over others. However, it is unclear how the account is supposed to go, given that attention must be driven by expected precisions.

To illustrate the problem that arises for the PC account, take the case of searching for one's keys. What are the relevant precision expectations driving attention? They cannot be spatial—one doesn't have high expected precision for any particular spatial region (beyond a few general expectations, such as that one's lost keys typically won't be found hanging from the ceiling). They are also not happily described as proprioceptive prediction errors—while it is true that some cases of endogenous attention are well described in terms of active inference, feature-based attention needn't involve any physical action on the part of the agent. Feature-based attention operates independently of spatial attention, and can "highlight" a given feature regardless of where it appears in the visual field (Carrasco 2011: 1507).

The relevant precision expectations must then be for perceptual prediction errors. In accordance with this, Hohwy (2015) explains feature-based attention as follows: "When the system endogenously attends to an as yet unseen feature, the precision of prediction error for that feature is expected to be high, which causes increased gain for that prediction error."

However, the issue in this case is that the visual system must first locate the relevant feature in order to apply the perceptual hypothesis, and only then can the gain be enhanced for that prediction error. One cannot enhance the gain on a prediction error that has not yet been generated, and the way to generate prediction error is to test one's hypothesis against the incoming sensory data. Moreover, the perceptual hypothesis must be applied selectively to the scene, because applying it indiscriminately to all objects would result in larger prediction errors being generated for all items that are not those items one is looking for, and as a consequence would result in an attentional pattern that is the inverse of what is seen in feature-based attention.

Bowman et al. (2013) raise a related worry for the PC account of endogenous attention:

> What makes attention so adaptive is that it can guide towards an object at an unpredictable location—simply on the basis of features. For example, we could ask the reader to find the nearest word printed in bold. Attention will typically shift to one of the headers, and indeed momentarily increase precision there, improving reading. But this makes precision weighting a *consequence* of attending. At least as interesting is the mechanism *enabling* stimulus selection in the first place. The brain has to first deploy attention before a precision advantage can be realized for that deployment.
>
> (207, emphasis original)

This criticism rests in part on a failure to appreciate the difference between precision expectations and the consequent precision weighting of prediction error. The former will determine the latter, and so high-precision expectations for the relevant feature are what guide attention to it. Nevertheless, there is still a question as to how precision expectations can guide attention when the spatial location of the object is not known. Contrasting this with spatial endogenous attention is instructive. In such cases precision expectations drive us to look in the right place, and so speed stimulus detection. Expecting precise prediction error from a perceptual hypothesis that has not yet been applied to a scene is different. In itself it does nothing to help select the relevant features. It is useful to consider Clark's response to Bowman et al.'s worry in detail. He writes, employing the example of looking for a four-leaf clover:

> The resolution of this puzzle lies, I suggest, in the potential assignment of precision-weighting at many different levels of the processing hierarchy. Feature-based attention corresponds, intuitively, to increasing the gain on the prediction error units associated with the identity or configuration of a stimulus (e.g. increasing the gain on units responding to the distinctive geometric pattern of a four-leaf clover). Boosting that response (by giving added weight to the relevant kind of sensory prediction error) should enhance detection of that featural cue. Once the cue is provisionally detected, the subject can fixate the right spatial region, now under conditions of "four-leaf-clover-there" expectation. Residual error is then amplified for that feature at that location, and high confidence in the presence of the four-leaf clover can (if you are lucky!) be obtained.
>
> (Clark 2013a: 238)

While this addresses the problem insofar as one accepts that "provisional detection" can guide attention, it raises the issue of how provisional detection is accomplished in PC. It too must function according to the same process of top-down hypothesis generation and revision, if PC is to be explanatorily complete. Perhaps provisional detection can be understood along the lines of "gist perception" (see Clark 2015: 163–4, Hohwy 2012: 2). Bar (2003) holds that perception is facilitated by the ability to first generate a prediction of the "gist" of the scene or object using low spatial frequency visual information that results in a basic-level categorization of the object's likely identity (see also Bar et al. 2001, Barrett and Bar 2009, Oliva and Torralba 2001, Schyns and Oliva 1994, Torralba and Oliva 2003). This then allows for the more fine-grained details to be filled in using the basic-level categorization as a guide. The idea here is that such basic-level categorization could guide selective application of the clover hypothesis, ensuring that it be applied only to objects that have the coarse-grained features of four-leaf clovers. This would then guide attention to the relevant spatial locations, privileging perceptual processing of these areas.

To be sure, the range of visual features that can be provisionally detected with this sort of gist perception is limited to what can be extracted from low spatial frequency information, namely, shapes and configurations (and so will exclude features like colors). Even with respect to these types of features, however, such a proposal is only a solution if the basic-level categorization itself is the result of PC, and here it is unclear as to whether the "gist" is constructed using the PC hierarchical framework. It certainly does not rely on high-level hypotheses such as "clover." Constructing the gist of a scene or object would rather be reliant on lower-level properties such as shape. It is then a further question whether such properties are detected in a feedforward model inconsistent with PC, or predicted in a feedback model consistent with PC. For example, rather than construing gist perception as an essential component of object recognition, Bar (2003) takes it to merely facilitate the process. Gist perception only reduces the time it takes to categorize an object, except perhaps in cases where the object is highly occluded or camouflaged. Bar (2003: 7) points to studies of patients with frontal cortex lesions as support for the ability of recognition to take place in a wholly bottom-up manner. Given Bar's own commitment to bottom-up processing it remains to be seen whether or not gist perception provides a solution to the problem at hand.

We can make the same point by examining more closely some other ways of modeling the top-down influence of expectations on visual search.[5] A central notion in many accounts of visual search is that of topographically organized "feature maps," which consist of neurons that respond to the same feature at different locations across the visual field (Humphreys and Mavritsaki 2012: 59). Activity in a given feature map indicates the presence of that feature at a particular location in the visual field. The activity of different feature maps is subsequently integrated in a central location map.

Given this setup, expectations about what some desired target looks like can guide the search by modulating the activity of different feature maps. In Wolfe's (1994) Guided Search model, for example, expectation for the particular features possessed by a target (e.g., its particular shape and color) pre-activates the corresponding feature maps. Thanks to this selective boosting, when the activity of various feature maps are integrated together in the central location map, spatial locations occupied by stimuli possessing the sought-after features exhibit heightened activity. This allows the spatial locations to be ordered in terms of their salience to the search, thus enabling an effective means for selecting the location of the desired target. In addition to this type of *target pre-activation*, there is evidence that similar effects are achieved by means of *non-target suppression*, where the activity of the feature maps associated with non-targets are selectively inhibited (Watson and Humphreys 1997).

On the face of it, the PC model of feature-based attention seems to fit neatly with the above account of visual search. In the case of the PC model, the expectation for features of the target is formulated in terms of an expectation that the precision of the prediction errors for those features is going to be high. What is more, preferentially increasing the gain on that prediction error leads to a "pre-emptive down-weighting of … prediction errors [associated with other non-target features]" (Hohwy 2015). It appears then that the PC model provides an economical way of capturing target pre-activation as well as non-target suppression.

However, in the Guided Search model sketched above, the feedforward information relayed from a given feature map contains information about the presence or absence of the feature at different locations across the visual field; the information does *not* merely consist of an "error response" to a prior prediction. Consequently, proponents of PC may be required to go beyond the talk of increasing the gain on the signals regarding the target features; they should explain the prior predictions that generate the signals (now seen as prediction error), the particular type of prediction error at issue, and so on.[6] Otherwise, PC must provide a clear explanation of feature-based attention without the assistance of tools that rely on a conception of bottom-up signals that extend beyond prediction errors.

3. The Precision Weighting of Prediction Error Is too Broad

Though postsynaptic gain is ubiquitous in the brain, attention is not synonymous with all such gain. It manifests only in the gain on prediction error, and not, for instance, with respect to gain on top-down predictions. Thus, it avoids the criticism that it is too broad a phenomenon to account for attention's selective functional role (Feldman and Friston 2010: 18). However, even with this constraint in place, the precision weighting of prediction error may be too broad to be synonymous with attention. The precision weighting of prediction error has been assigned a dizzying number of roles. Along with explaining how we switch between model-based and model-free cognitive strategies, and its potential role in planning for the future (see section 2.3), it has also been invoked to explain our sense of agency, and how we understand the actions of others (Clark 2013b). To understand the actions of others, we observe a person's behavior and then simulate it as if we were the ones performing the action, generating hypotheses as to the intention behind the behavior. By assigning an extremely low (imprecise) weight to proprioceptive prediction errors generated by lower levels of the hierarchy, we prevent ourselves from actually performing the actions of our mental simulations. Our sense of agency arises in cases where proprioceptive prediction error is highly weighted but resolved by our top-down predictions.

In all such instances PC holds that shifts in precision weighting of prediction errors are constitutive of attentional shifts. Therefore, in each case we may ask whether the hypothesized attentional patterns correspond to our actual attentional behavior. Here we will focus on the case of multimodal integration, which involves the integration of information from the different sensory modalities—sight, sound, smell, taste, touch, and the sense of one's body in space (proprioception)—to form a coherent and unified percept. We perceive the smell as coming from the lillies and the surf of the ocean, the sound as being produced by the seagull, the flavor as deriving from the mango we are eating. Our expected precisions for these modalities may differ according to context—in dim lighting we may weight our auditory prediction errors more heavily, and in a noisy crowd we may be more confident in the deliverances of our visual prediction errors. We also may have more stable differential precision weightings, such as expecting visual information to provide more reliable fine-grained locational information than sound.

Differential precision weighting plays a central role in explaining crossmodal illusions that result from multimodal integration, such as the McGurk effect (McGurk and MacDonald 1976). In this illusion, participants watch a video of a person's lips moving as if making a certain sound together with audio of a different sound. Participants usually report hearing a third sound that can be understood as a sort of "compromise" between the visual and auditory information. For example, when the auditory information is "ba" and the visual information is "ga" then participants report hearing "da." This can be explained on PC by positing that we expect visual information pertaining to lip movements to be relatively precise in comparison to auditory information, all else equal, and so this information "overrides" the auditory signal to some degree in order to produce the illusion (e.g., Miller and Clark 2018: 2566–7). This precision expectation can also shift with context—when the auditory information is degraded, then the illusion is more pronounced (Massaro and Cohen 2000). On PC this would be explained by a lower relative precision weighting of auditory prediction error over visual prediction error in this context.

With a basic understanding of the account, we can turn to the question of whether the precision weighting of modality-specific information can be identified with attention. We argue that it cannot. This is most evident in the case of crossmodal illusions such as the McGurk effect. On PC, a lower precision weighting of the prediction error generated by audition should correspond to decreased auditory attention. But the illusion is not accompanied by a decrease in the apparent volume of the speaker. Instead we mishear the sound, at its original volume. Precision weighting of prediction error in this case seems to work not by shifting our attention away from the relevant sensory modality, but rather by causing us to alter the content of that modality.

Other crossmodal illusions pose the same challenge. In the case of the rubber hand illusion (Botvinick and Cohen 1998), the participant's arm is kept out of view and a rubber arm is placed on the table in front of her, in the same place her own hand might reasonably be located given her position. An experimenter then touches, tickles, or pokes the rubber arm, while also simultaneously performing the same actions on the participant's hidden arm. The result is usually that the participant begins to experience an illusion that the rubber arm is her own, and will respond to threats to the rubber hand in the same way as she responds to threats to her real hand (Ehrssen et al. 2007).

PC provides an explanation of the illusion in terms of the higher relative precision weighting of visual and tactile prediction errors over proprioceptive prediction errors (Hohwy 2013: 107).[7] Though we initially begin with the sense of our arm as being situated in its actual location, out of sight, the visual information of the rubber arm being manipulated, along with the sensations in our own arm or hand, causes us to shift our sense of where our arm is located in space. Again, this illusion is impervious to our knowledge that the arm is not really our own, but made of rubber. This sort of illusion is also problematic insofar as the incongruous visual and proprioceptive hypotheses are resolved not by diminishing our attention to where our arm is located in space, but by changing our sense of location. In fact, we may pay more attention than ever to our sense of bodily location just because the illusion is so bizarre.

But perhaps there is a PC story here about how, even though proprioceptive prediction error is assigned a smaller precision weight at lower levels of the hierarchy, attention is only allocated at higher levels of the hierarchy, to multimodal hypotheses. This is consistent with PC's proposal of how our perceptual systems solve the problem of binding inputs from disparate sensory modalities to form a unified percept—they propose that instead of trying to bind together disparate sensory inputs, the top-down hypotheses generated on PC already presuppose that the inputs are bound (see Hohwy 2013: ch. 5). They are already multimodal.

While this seems a reasonable solution, note that the precision weighting of prediction error is no longer synonymous with attention at lower levels of the hierarchy. The solution does not address our criticism that the precision weighting of prediction error is broader than attentional phenomena. The examples provided here suggest that, at least at lower levels of the generative model's hierarchy, precision-weighted prediction error cannot be identified with attention.

4. Conclusion

In this chapter we have argued that attention cannot be straightforwardly equated with the optimization of expected precisions, as some PC theorists have suggested. It is simultaneously too narrow and too broad a concept. The optimization of precision expectations is too narrow in that it may fail to account for a variety of attentional phenomena. Affect-biased attention suggests that attention to affectively salient objects can occur even in the absence of prediction error that is expected to be precise. PC faces difficulties in explaining voluntary attention because while arbitrary decisions to attend to things do result in an increase in precision, such decisions cannot meaningfully said to be driven by precision expectations; it is equally true of all possible objects of attention that precision will be enhanced by deciding to attend. Whether or not PC can explain intellectual attention will depend on the details of the account, but we suspect that shifts in intellectual attention are also sometimes affectively driven, and so also not fully accommodated by precision expectations. Feature-based attention may not be accounted for by PC because it appears to rely in part on bottom-up processes not explainable in terms of prediction error.

Finally, we have argued that the optimization of precision expectations is too broad a phenomenon to be identified with attention because such optimization is crucially involved in multimodal integration. PC theory explains crossmodal illusions in terms of differential precision weightings, but this does not translate into attentional phenomena as one would expect on PC.

As a whole the criticisms we have provided here suggest that PC needs to go beyond the precision weighting of prediction error to accommodate the full range of attentional phenomena, thus hampering its claims to completeness. At minimum, the challenges we have raised here are an opportunity for PC theorists to clarify important details of their account.[8]

Notes

1 For a review of different PC algorithms, see Spratling (2017).
2 See Aitchison and Lengyel (2017) for discussion of how these concepts come apart. PC is also often combined with the free energy principle. Free energy is an information-theoretic measure; it bounds the evidence for our internal generative models of sensory data (MacKay 1995). The free energy principle is that all adaptive changes in the brain will minimize free energy, where such changes range from those that occur on evolutionary time scales to those happening in real-time (Friston 2009). Under some simplifying assumptions, to minimize free energy is to minimize prediction error (Friston 2010).
3 See Rauss and Pourtois (2013) for discussion of how to understand "top-down" and "bottom-up" in PC.
4 We might retrieve the answer "Cervantes" or "Pierre Menard," depending on one's view of the ontological status of literary works.
5 See Humphreys and Mavritsaki (2012) for an informative summary of different models of visual search.
6 For instance, neurophysiological studies often distinguish between two types of prediction errors pertaining to a feature or an object: positive prediction error corresponds to the occurrence of an unexpected instance of that feature or object. Negative prediction error pertains to the unexpected omission of that feature or object (Egner et al. 2010). Clearly, it makes a difference whether the type of prediction error generated during the search is due to occurrence or omission of the feature and very different consequences would follow from increasing the gain on each of these types.
7 For an understanding of how interoceptive inference might also play a role in the illusion, see Suzuki et al. (2013).
8 Thanks to Jakob Hohwy, Carolyn Dicey-Jennings, and David Barack for their comments on earlier versions of this chapter, as well as the audience at the 2015 Society for Philosophy and Psychology annual conference and at the 2015 Minds Online conference. We would also like to acknowledge the financial support of the Summer Seminar for Neuroscience and Philosophy at Duke University, through which a portion of this research was conducted.

References

Anderson, B. A. (2013), "A Value-Driven Mechanism of Attentional Selection," *Journal of Vision*, 13 (3): 1–16.

Anderson, B. A. and Yantis, S. (2013), "Persistence of Value-driven Attentional Capture," *Journal of Experimental Psychology: Human Perception and Performance*, 39 (1): 6.

Anderson, B. A., Laurent, P. A. and Yantis, S. (2011), "Learned Value Magnifies Salience-Based Attentional Capture," *PloS one*, 6 (11): e27926.

Anderson, B. A., Laurent, P. A. and Yantis, S. (2012), "Generalization of Value-based Attentional Priority," *Visual Cognition*, 20 (6): 647–58.

Anderson, B. A., Laurent, P. A. and Yantis, S. (2013), "Reward Predictions Bias Attentional Selection," *Frontiers in Human Neuroscience*, 7: 262.

Aitchison, L. and Lengyel, M. (2017), "With or Without You: Predictive Coding and Bayesian Inference in the Brain," *Current Opinion in Neurobiology*, 46: 219–27.

Awh, E., Belopolsky, A. V. and Theeuwes, J. (2012), "Top-Down versus Bottom-Up Attentional Control: A Failed Theoretical Dichotomy," *Trends in Cognitive Sciences*, 16 (8): 437–43.
Bar, M. (2003), "A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition," *Journal of Cognitive Neuroscience*, 15 (4): 600–9.
Bar, M., Tootell, R., Schacter, D. L., Doug, G. N., Fischl, B., Mendola, J. D., Rosen, B. R. and Dale, A. M. (2001), "Cortical Mechanisms Specific to Explicit Visual Object Recognition," *Neuron*, 29 (2): 529–35.
Barrett, L. F. and Bar, M. (2009), "See It with Feeling: Affective Predictions during Object Perception," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364 (1521): 1325–34.
Botvinick, M. and Cohen, J. (1998), "Rubber Hands 'Feel' Touch That Eyes See," *Nature*, 391 (6669): 756.
Bowman, H., Filetti, M., Wyble, B. and Olivers, C. (2013), "Attention Is More than Prediction Precision," *The Behavioral and Brain Sciences*, 36 (3): 206–8.
Carrasco, M. (2011), "Visual Attention: The Past 25 Years," *Vision Research*, 51 (13): 1484–525.
Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibánez, A., Owen, A. M. and Bekinschtein, T. A. (2013), "Expectation and Attention in Hierarchical Auditory Prediction," *Journal of Neuroscience*, 33 (27): 11194–205.
Christoff, K. (2012), "Undirected Thought: Neural Determinants and Correlates," *Brain Research*, 1428: 51–9.
Clark, A. (2013a), "Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science," *Behavioral and Brain Sciences*, 36 (3): 181–204.
Clark, A. (2013b), "The Many Faces of Precision (Replies to Commentaries on 'Whatever Next? Neural Prediction, Situated Agents, and the Future of Cognitive Science')," *Frontiers in Psychology*, 4: 270.
Clark, A. (2015), *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*, New York: Oxford University Press.
Clark, A. (2017), "Predictions, Precision, and Agentive Attention," *Consciousness and cognition*, 56: 115–19.
Colombo, M. and Wright, C. (2017), "Explanatory Pluralism: An Unrewarding Prediction Error for Free Energy Theorists," *Brain and Cognition*, 112: 3–12.
Daw, N. D., Niv, Y. and Dayan, P. (2005), "Uncertainty-based Competition between Prefrontal and Dorsolateral Striatal Systems for Behavioral Control," *Nature Neuroscience*, 8 (12): 1704.
Della Libera, C. and Chelazzi, L. (2009), "Learning to Attend and to Ignore Is a Matter of Gains and Losses," *Psychological Science*, 20 (6): 778–84.
Egner, T., Monti, J. M. and Summerfield, C. (2010), "Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream," *Journal of Neuroscience*, 30 (49): 16601–16608.
Egner, T., Monti, J., Trittschuh, E. H., Wieneke, C. A., Hirsch, J. and Mesulam, M.-M. (2008), "Neural Integration of Top-Down Spatial and Feature-based Information in Visual Search," *Journal of Neuroscience*, 28 (24): 6141–51.
Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J. and Passingham, R. E. (2007), "Threatening a Rubber Hand That You Feel Is Yours Elicits a Cortical Anxiety Response," *Proceedings of the National Academy of Sciences*, 104 (23): 9828–33.
Feldman, H. and Friston, K. J. (2010), "Attention, Uncertainty, and Free-energy," *Frontiers in Human Neuroscience*, 4: 215.

Fortney, M. (2019), "Conceptualizing Intellectual Attention," *Theory & Psychology*, 29 (6): 775–88.
Friston, K., Mattout, J. and Kilner, J. (2011), "Action Understanding and Active Inference," *Biological Cybernetic*, 104 (1–2): 137–60.
Friston, K. J. (2008), "Hierarchical Models in the Brain," *PLoS Computational Biology*, 4 (11): e1000211.
Friston, K. J. (2009), "The Free-Energy Principle: A Rough Guide to the Brain?," *Trends in Cognitive Sciences*, 13 (7): 293–301.
Friston, K. J. (2010), "The Free-Energy Principle: A Unified Brain Theory?," *Nature Reviews Neuroscience*, 11 (2): 127.
Friston, K. J., Daunizeau, J., Kilner, J. and Kiebel, S. J. (2010), "Action and Behavior: A Free-Energy Formulation," *Biological Cybernetics*, 102 (3): 227–60.
Gershman, S. J. and Daw, N. D. (2012), "Perception, Action and Utility: The Tangled Skein," *Principles of Brain Dynamics: Global State Interactions*: 293–312. doi.org/10.7551/mitpress/9108.003.0015.
Helmholtz, H. V. (2005), *Treatise on Physiological Optics*, Mineola: Dover.
Hickey, C., Chelazzi, L. and Theeuwes, J. (2010), "Reward Changes Salience in Human Vision Via the Anterior Cingulate," *Journal of Neuroscience*, 30 (33): 11096–11103.
Hobson, J. A. and Friston, K. J. (2012), "Waking and Dreaming Consciousness: Neurobiological and Functional Considerations," *Progress in Neurobiology*, 98 (1): 82–98.
Hohwy, J. (2012), "Attention and Conscious Perception in the Hypothesis Testing Brain," *Frontiers in Psychology*, 3: 96.
Hohwy, J. (2013), *The Predictive Mind*, New York: Oxford University Press.
Hohwy, J. (2015), "Comments on Ransom and Fazelpour's Three Problems for the Predictive Coding Theory of Attention," *Minds Online*. http://mindsonline.philosophyofbrains.com/2015/session4/three-problems-for-the-predictive-coding-theory-of-attention/#comment–284.
Hohwy, J., Roepstorff, A. and Friston, K. J. (2008), "Predictive Coding Explains Binocular Rivalry: An Epistemological Review," *Cognition*, 108 (3): 687–701.
Huang, Y. and Rao, R. (2011), "Predictive Coding," *Wiley Interdisciplinary Reviews: Cognitive Science*, 2 (5): 580–93.
Humphreys, G. W. and Mavritsaki, E. (2012), "Models of Visual Search: From Abstract Function to Biological Constraint" in Michael I. Posner (ed.), *Cognitive Neuroscience of Attention*, New York: The Guilford Press.
Jiang, J., Summerfield, C. and Egner, T. (2013.), "Attention Sharpens the Distinction Between Expected and Unexpected Percepts in the Visual Brain," *Journal of Neuroscience*, 33 (47): 18438–47.
Kok, P., Rahnev, D., Jehee, J., Lau, H. C. and. De Lange, F. P. (2011), "Attention Reverses the Effect of Prediction in Silencing Sensory Signals," *Cerebral Cortex*, 22 (9): 2197–206.
Körding, K. P. and Wolpert, Daniel M. (2006), "Bayesian Decision Theory in Sensorimotor Control," *Trends in Cognitive Sciences*, 10 (7): 319–26.
MacKay, D. J. (1995), "Free Energy Minimisation Algorithm for Decoding and Cryptanalysis," *Electronics Letters*, 31 (6): 446–7.
Markovic, J., Anderson, A. K. and Todd, R. M. (2014), "Tuning to the Significant: Neural and Genetic Processes Underlying Affective Enhancement of Visual Perception and Memory," *Behavioural Brain Research*, 259: 229–41.

Massaro, D. W. and Cohen, M. M. (2000), "Tests of Auditory–Visual Integration Efficiency within the Framework of the Fuzzy Logical Model of Perception," *The Journal of the Acoustical Society of America*, 108 (2): 784–9.

Mather, M. and Sutherland, M. R. (2011), "Arousal-biased Competition in Perception and Memory," *Perspectives on Psychological Science*, 6 (2): 114–33.

McGurk, H. and MacDonald, J. (1976), "Hearing Lips and Seeing Voices," *Nature*, 264 (5588): 746.

Miller, M. and Clark, A. (2018), "Happily Entangled: Prediction, Emotion, and the Embodied Mind," *Synthese*, 195 (6): 2559–75.

Mole, C. (2011), *Attention Is Cognitive Unison: An Essay in Philosophical* Psychology, New York: Oxford University Press.

Niu, Y., Todd, R. and Anderson, A. K. (2012), "Affective Salience Can Reverse the Effects of Stimulus-Driven Salience on Eye Movements in Complex Scenes," *Frontiers in Psychology*, 3: 336.

Oliva, A. and Torralba, A. (2001), "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," *International Journal of Computer Vision*, 42 (3): 145–75.

Posner, M. I. (1980), "Orienting of Attention," *Quarterly Journal of Experimental Psychology*, 32 (1): 3–25.

Posner, M. I. (ed.) (2011), *Cognitive Neuroscience of Attention*, New York: Guilford Press.

Ransom, M., Fazelpour, S. and Mole, C. (2017), "Attention in the Predictive Mind," *Consciousness and Cognition*, 47: 99–112.

Ransom, M., Fazelpour, S., Markovic, J., Kryklywy, J., Thompson, E. T. and Todd, R. M. (2020), "Affect-Biased Attention and Predictive Processing," *Cognition*, 203: 104370.

Rao, R. and Ballard, D. H. (2005), "Probabilistic Models of Attention Based on Iconic Representations and Predictive Coding" in Laurent Itti and Geraint Rees (eds.), *Neurobiology of Attention*, Cambridge: Academic Press.

Rauss, K. and Pourtois, G. (2013), "What Is Bottom-up and What Is Top-Down in Predictive Coding?," *Frontiers in Psychology*, 4: 276.

Rolls, E. T. (2000), "On the Brain and Emotion," *Behavioral and Brain Sciences*, 23 (2): 219–28.

Rossi, A. F. and Paradiso, M. A. (1995), "Feature-specific Effects of Selective Visual Attention," *Vision Research*, 35 (5): 621–34.

Schyns, P. G. and Oliva, A. (1994), "From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dependent Scene Recognition," *Psychological Science*, 5 (4): 195–200.

Seth, A. K. and Critchley, H. D. (2013), "Extending Predictive Processing to the Body: Emotion as Interoceptive Inference," *Behavioral and Brain Science*, 36 (3): 227–8.

Seth, A. K., Suzuki, K. and Critchley, H. D. (2012), "An Interoceptive Predictive Coding Model of Conscious Presence," *Frontiers in Psychology*, 2: 395.

Shomstein, S. and Johnson, J. (2013), "Shaping Attention with Reward: Effects of Reward on Space-and Object-based Selection," *Psychological* Science, 24 (12): 2369–78.

Spratling, M. W. (2008), "Predictive Coding as a Model of Biased Competition in Visual Attention," *Vision Research*, 48 (12): 1391–408.

Spratling, M. W. (2017), "A Review of Predictive Coding Algorithms," *Brain and Cognition*, 112: 92–7.

Srinivasan, M. V., Laughlin, S. B. and Dubs, A. (1982), "Predictive Coding: A Fresh View of Inhibition in the Retina," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 216 (1205): 427–59.

Summerfield, C. and Egner, T. (2009), "Expectation (and Attention) in Visual Cognition," *Trends in Cognitive Sciences*, 13 (9): 403–9.

Suzuki, K., Garfinkel, S. N., Critchley, H. D. and Seth, A. K. (2013), "Multisensory Integration Across Exteroceptive and Interoceptive Domains Modulates Self-experience in the Rubber-hand Illusion," *Neuropsychologia*, 51 (13): 2909–17.

Todd, R., Cunningham, W. A., Anderson, A. K. and Thompson, E. (2012), "Affect-biased Attention as Emotion Regulation," *Trends in Cognitive Sciences*, 16 (7): 365–72.

Todd, R. M. and Manaligod, M. (2018), "Implicit Guidance of Attention: The Priority State Space Framework," *Cortex*, 102: 121–38.

Torralba, A. and Oliva, A. (2003), "Statistics of Natural Image Categories," *Network: Computation in Neural Systems*, 14 (3): 391–412.

Vuilleumier, P. (2015), "Affective and Motivational Control of Vision," *Current Opinion in Neurology*, 28 (1): 29–35.

Watson, D. G. and Humphreys, G. W. (1997), "Visual Marking: Prioritizing Selection for New Objects by Top-Down Attentional Inhibition of Old Objects," *Psychological Review*, 104 (1): 90.

Watzl, S. (2017), *Structuring Mind: The Nature of Attention and How It Shapes Consciousness*, New York: Oxford University Press.

Wilkinson, S. (2014), "Accounting for the Phenomenology and Varieties of Auditory Verbal Hallucination within a Predictive Processing Framework," *Consciousness and Cognition*, 30: 142–55.

Wolfe, J. M. (1994), "Guided Search 2.0: A Revised Model of Visual Search," *Psychonomic Bulletin &* Review, 1 (2): 202–38.

Wu, W. (2014), *Attention*, London: Routledge.