Cognition 203 (2020) 104370

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Affect-biased attention and predictive processing

Madeleine Ransom[d,*], Sina Fazelpour[e], Jelena Markovic[a], James Kryklywy[b], Evan T. Thompson[a], Rebecca M. Todd[b,c]

[a] University of British Columbia, Department of Philosophy, Canada
[b] University of British Columbia, Department of Psychology, Canada
[c] University of British Columbia, Centre for Brain Health, Canada
[d] Indiana University Bloomington, Department of Cognitive Science, USA
[e] Carnegie Mellon University, Department of Philosophy, USA

ARTICLE INFO

Keywords:
Predictive processing
Affective salience
Attention
Free energy
Bayesian brain
Emotion

ABSTRACT

In this paper we argue that predictive processing (PP) theory cannot account for the phenomenon of affect-biased attention – prioritized attention to stimuli that are affectively salient because of their associations with reward or punishment. Specifically, the PP hypothesis that selective attention can be analyzed in terms of the optimization of precision expectations cannot accommodate affect-biased attention; affectively salient stimuli can capture our attention even when precision expectations are low. We review the prospects of three recent attempts to accommodate affect with tools internal to PP theory: Miller and Clark's (2018) embodied inference; Seth's (2013) interoceptive inference; and Joffily and Coricelli's (2013) rate of change of free energy. In each case we argue that the account does not resolve the challenge from affect-biased attention. For this reason, we conclude that prediction error minimization is not sufficient to explain all mental phenomena, contrary to the claim that the PP framework provides a unified theory of all mental phenomena or the brain's cognitive functioning. Nevertheless, we suggest that empirical investigation of the interaction between affective salience and precision expectations should prove helpful in understanding the limits of PP theory, and may provide new directions for the application of a Bayesian perspective to perception.

1. Introduction: explaining the challenge to PP theory of attention

Suppose you walk your dog uneventfully every day past a house on the corner of your block. One morning, however, a large Doberman rushes to the fence, barking and snapping. You jump backwards and for a moment you fear for your life. From this day forward, you give this house a bit of extra attention when you walk past, your eyes always searching the fence for signs of the Doberman, though it is seldom in fact in the yard. What explains your change in behavior? A common sense approach would suggest the answer is rather obvious - you do not like to be startled, and so you selectively attend to that fence over other aspects of the environment just in case the dog is there, irrespective of the probability.

As simple as that seems, cases such as these appear to pose a significant challenge to predictive processing (PP) models of attention, and therefore to the prospects of PP as a unified theory of all mental functioning (Clark, 2013, 2015; Feldman & Friston, 2010; Hohwy, 2012, 2013). Specifically, the PP account of attention as *the optimization of precision expectations* holds that attention is allocated on the basis of selecting those signals expected to be most reliably informative, or 'highly precise,' where a signal's precision is the inverse of its variance. However, this does not currently capture the phenomenon of affect-biased attention – attention to stimuli that are affectively salient because of their associations with reward or punishment. Such stimuli can drive attention in spite of correspondingly low precision expectations, suggesting that the PP analysis of attention is inadequate. This is in contrast to Bayesian decision theory (BDT), which may have the resources to accommodate affect-biased attention, though actual models are lacking and attentional phenomena have been largely unexplored. Nevertheless, understanding where PP goes wrong is instructive for future BDT models of affect-biased attention.

1.1. Predictive processing

Predictive Processing (PP) purports to provide a unified account of all mental functioning, where the brain's single overarching task is to

* Corresponding author at: Indiana University Bloomington, Department of Cognitive Science, 819 Eigenmann, 1900 E. 10th St., IN 47406-7512, USA.
E-mail address: madeleineransom@gmail.com (M. Ransom).

https://doi.org/10.1016/j.cognition.2020.104370
Received 16 May 2019; Received in revised form 22 May 2020; Accepted 3 June 2020

minimize surprise. It is a synthesis of several theories that will be elaborated on below: Bayesian decision theory, the Free Energy Principle, and predictive coding.

1.1.1. Bayesian decision theory

Against the view that perceivers are passive consumers of sensory input, Helmholtz (2005) maintained that we actively – though unconsciously – infer the causes of this input. His theory was developed in part to solve an underdetermination problem: the sensory input we receive is impoverished in that it is consistent with having been caused by many different possible states of the world around us. By actively predicting the hidden causes of our sensory input, and testing these predictions against further sensory input, we can come to more accurately perceive the world.

A strand of research in perceptual psychology builds on this proposal by invoking Bayesian decision theory (BDT), which provides a mathematical framework for decision-making under uncertainty (Knill & Richards, 1996). To solve the underdetermination problem, our perceptual systems possess an internal model or models of the external world, which generate hypotheses as to the worldly cause of the sensory input.

In Bayesian inference, the agent – or some cognitive system therein – is endowed with a hypothesis space, H, which is the totality of hypotheses for a given sensory cause. When a piece of sensory input, e, is received, the subjective probability attached to each hypothesis, $h \in H$, is updated according to Bayes's Rule.

$$\mathrm{p(h \mid e)} = \frac{\mathrm{p(h)} \times \mathrm{p(e \mid h)}}{\mathrm{p(e)}}$$

where $p(h)$ is the prior probability of hypothesis, h and $p(e|h)$ is the likelihood of observing e if the hypothesis were true. The denominator $p(e)$ is the marginal probability of evidence and is calculated as $p(e) = \Sigma_{h \in H} p(h, e)$. Bayesian updating involves the assignment of the posterior probability, $p(h|e)$, as the new prior. It is this process of hypothesis generation and manipulation that allows the agent or cognitive system to quantify and update its uncertainties concerning the causes of ambiguous sensory input.

In BDT, the agent or cognitive system uses Bayesian updating to make optimal choices. That is, the agent can compute posteriors concerning the likely states of the world, if, given the current sensory input, e, the agent were to act on some policy, and then use these posteriors to choose the policy that, in light of e, maximizes the expected utility (or minimizes loss). Unfortunately, in practice, calculating $p(e)$, which is needed for inferring the posterior, is rather difficult and can quickly become intractable for reasonably complex hypothesis spaces. The free energy principle provides one method of approximating Bayesian inference in the brain.

1.1.2. The free energy principle

Simply put, the Free Energy Principle states that all adaptive changes in brain functioning, from those on evolutionary timescales to those occurring in real-time, can be explained in relation to the task of minimizing free energy (2009; see also Friston, Kilner, & Harrison, 2006). Free energy is a concept used in statistics to approximate inference via variational Bayes. The free energy principle thus applies a particular approximation of BDT, namely variational inference via free energy minimization, to explain all adaptive behavior.

In order to see how this fits in with the discussion of Bayesian inference thus far, consider that, due to the difficulty of computing $p(e)$, directly selecting the hypothesis that would maximize the posterior probability is not a feasible objective. Progress could be made, however, if we were to make additional assumptions about the phenomenon of interest. Let us assume, for example, that there is a family of distributions, Q, that (ideally) includes the true conditional distribution $p(h|e)$ that interests us. Instead of searching over all possible posterior distributions, we can then approximate $p(h|e)$ by finding a member of Q that is, in some sense, closest to $p(h|e)$. Using the Kullback-Leibler divergence as our measure of distance, our task is thus to find $q^* \in Q$ such that

$$\mathrm{q^* = argmin\ D_{KL}(q(h) \| p(h|e))}$$

where $D_{KL}(q(h)||p(h \mid e)) = \mathbb{E}[\log q(h)] - \mathbb{E}[\log p(h|e)]$, with both expectations defined with respect to $q(h)$.

By itself, this does not spare us from the difficulty of computing $p(e)$. For, expanding the divergence term, it is clear that:

$$\begin{aligned} D_{KL}(q(h)\|p(h \mid e)) &= \mathbb{E}[\log q(h)] - \mathbb{E}[\log p(h|e)] \\ &= \mathbb{E}[\log q(h)] - \mathbb{E}[\log p(he)] + \mathbb{E}[\log p(e)] \\ &= \mathbb{E}[\log q(h)] - \mathbb{E}[\log p(he)] + \log p(e) \end{aligned}$$

The last step follows from the fact that $\log p(e)$ is constant with respect to $q(h)$. The challenge of computing $p(e)$ thus remains, now in the form of the task of calculating $\log p(e)$ – the negative of an information-theoretic quantity known as surprise. Nonetheless, we can approach the minimization task *indirectly* by re-formulating the problem in terms of free energy, $\mathscr{F}$—an information-theoretic measure that provides an upper bound on surprise (MacKay, 1995). Specifically, we can calculate $\mathscr{F}$ as

$$\mathscr{F} = D_{KL}(q(h)\|p(h \mid e)) - \log p(e)$$

Notice that since $\log p(e)$ is constant with respect to $q(h)$, minimizing free energy is equivalent to minimizing the KL divergence (Blei, Kucukelbir, & McAuliffe, 2017). This equivalence is useful because, in cases where the objective of directly minimizing the divergence is intractable, the equivalence offers an approximation strategy. For in such cases, the objective of minimizing free energy—computable as $\mathbb{E}\left[\log \frac{q(h)}{p(h,e)}\right]$—is something that we *can* aspired towards, if further assumptions are made to simplify and restrict the problem space.

These simplifying assumptions might include restrictions on the family of distributions, assumptions about the organization of the internal generative model, and so on (Gershman, 2019). For instance, theorists employing BDT often posit that the model contains a number of levels, organized hierarchically, with higher levels of the model generating hypotheses about more abstract and slower external-world regularities as compared to the lower levels (Lee & Mumford, 2003; Rohe & Noppeney, 2015).

It is this hierarchical version of BDT (formulated in terms of free energy) that PP adopts. According to this view, the communication between the levels is limited, with each level communicating only with the level directly above or below it in the hierarchy. Hypotheses at higher levels of the hierarchy involve variables that lead to, or generate, more specific hypotheses at lower levels. Thus, higher-level hypotheses either directly or indirectly constrain the lower-level hypotheses, though they are in turn revised in light of the success or failure of the lower-level hypotheses. For example, a higher-level hypothesis might concern an object's identity or overall shape, whereas a lower level hypothesis might concern itself with specific characteristics of the object, such as component shapes or textures.

With this hierarchy in place, the PP view is almost complete. The last theoretical commitment of PP is predictive coding.

1.1.3. Predictive coding

Predictive coding (not to be confused with PP, which involves more substantive commitments) is a strategy for minimizing information transmission whereby the difference between a prediction and an input is represented instead of representing the input directly, and only this difference, or prediction error is transmitted (Elias, 1955). Proposed as a computational model of visual processing, it has been successful in accounting for extra-classical receptive field effects (Rao & Ballard, 1999). In human perceivers, retinal ganglion cells have been hypothesized to engage in predictive coding, where neural circuits predict likely image characteristics of nearby spatial locations based on local image

characteristics, subtracting the predicted from actual values (Hosoya, Baccus, & Meister, 2005).

While predictive coding needn't invoke BDT (Aitchison & Lengyel, 2017), the view that we focus on in this paper – PP – does (Clark, 2013, 2015; Friston, 2008, 2009; Hohwy, 2012, 2013). It posits the hierarchical Bayesian model discussed in 1.1.2, with the added assumption that when there is a mismatch between hypothesis and input what gets relayed up the hierarchy to the level above it is merely prediction error – the data that has not been successfully predicted by the hypothesis. This in turn will lead to revisions of the predictions until prediction error is minimized. Because only error is passed up to the next level of the hierarchy, the system cuts down on the total amount of information transmitted.

This conservation of cognitive energy is a strong motivation for PP theorists to adopt predictive coding. Indeed, to say that we minimize free energy is equivalent to saying – given some simplifying assumptions – that we minimize prediction error (Bogacz, 2017). This, despite the fact that predictive coding and the free energy principle also come apart: those who endorse predictive coding needn't take on board the free energy principle, and the free energy principle may still prove correct even if predictive coding is not the means by which neural signalling works.

1.1.4. Summary

In summary, PP involves several theoretical commitments that go beyond extant BDT models – chiefly an adherence to predictive coding and to the free energy principle. Keeping these distinctions in mind is important in evaluating evidence for PP. Bayesian modelling in perceptual psychology has enjoyed considerable success in explaining various perceptual phenomena (Brainard & Gazzaniga, 2009; Ernst, 2010; Mamassian, Landy, & Maloney, 2002; Stone, 2011; Weiss, Simoncelli, & Adelson, 2002). However, this success does not vindicate PP. Indeed, there is some difficulty with pinning down falsifiable hypotheses that are specific to PP (Gershman, 2019). There is nevertheless some evidence for PP, predominantly in the perceptual domain (Bendixen, SanMiguel, & Schröger, 2012; Hohwy, Roepstorff, & Friston, 2008; Huang & Rao, 2011; Moreno-Bote, Knill, & Pouget, 2011; Stefanics, Kremláček, & Czigler, 2014; Summerfield & Koechlin, 2008; Todorovic & de Lange, 2012; Todorovic, van Ede, Maris, & de Lange, 2011; Wacongne, Changeux, & Dehaene, 2012) but see (Aitchison & Lengyel, 2017; Heeger, 2017). Even here, though, care must be taken because much of this evidence is for hierarchical predictive coding, and so should not be taken as a vindication in of itself for the free energy principle component of PP.

It is also important to note that the explanatory ambitions of PP are considerably broader than those of extant BDT models. While one might support the claim that some aspects of cognition involve an approximation of BDT, adopting the free energy principle offers PP theorists an ambitious unifying operating principle for the brain, with all our systems understood as dedicated to minimizing free energy. The principle – together with its other theoretical posits – purports to be able to parsimoniously explain not only perception, but all psychological phenomena. Much of the allure of such an account is thus that one need only posit a single type of mechanism for all kinds of mental activity, including action and attention (Hohwy, 2013, p.2).

1.2. Active inference

The PP theory of action, called 'active inference,' utilizes the same tools as perceptual inference. An internal, hierarchically structured generative model makes proprioceptive predictions updated according to Bayes' rule, where prediction error is the only feedforward input. Again, active inference is understood to be the minimization of proprioceptive prediction error. However, while in the case of perceptual inference we revise our predictions to fit the world, in the case of active inference we change the world to fit with our predictions (Adams, Shipp, & Friston, 2013; Clark, 2013, 2015; Friston, Daunizeau, Kilner, & Kiebel, 2010; Hohwy, 2013). My desire to drink a glass of water will involve generating a hypothesis that I am drinking the water, and minimizing prediction error will involve acting to make it the case that I

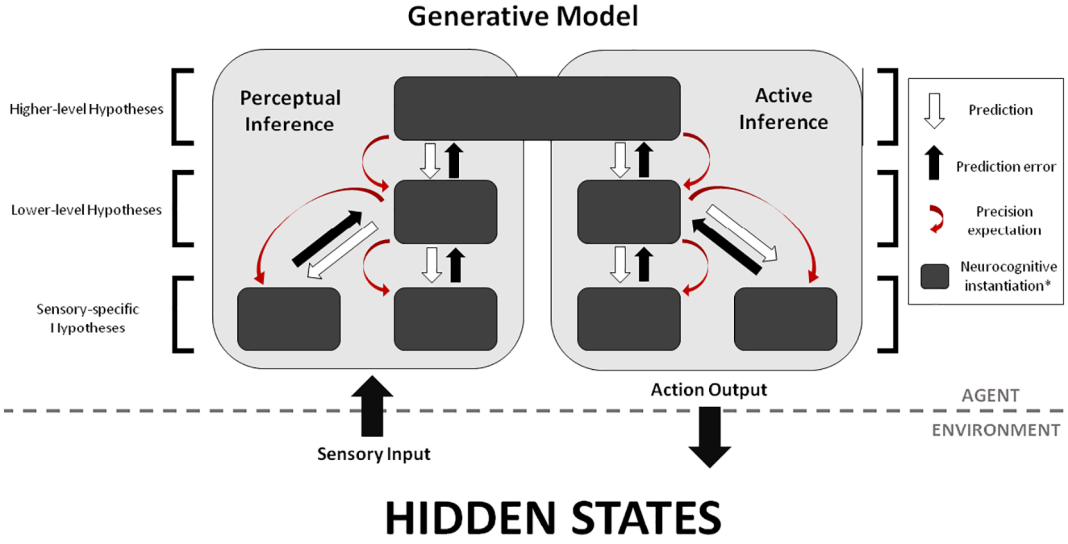

Fig. 1. Conceptual representation of the *Predictive Processing* (PP) account for combined actions and sensory stimuli. PP postulates a hierarchically structured* model that generates hypotheses concerning the hidden states of the world, with hypotheses at higher levels constraining those at lower levels. Higher levels of the hierarchy correspond to hypotheses concerning more abstract and slower regularities, and lower levels correspond to sensory and motor-specific hypotheses (perceptual and active inferences respectively; white arrows). Inconsistency between predictions and incoming sensory signals/motor expectation results in an error signal (prediction error; black arrows) being fed 'backwards' through the system, which forces a revision to the higher-level hypotheses until prediction errors are minimized. Prediction errors are weighted by their precision expectations (i.e., their reliability in representing the true state of the world; red/curved arrows), thus PP prioritizes the allocation of attention to sensory information (prediction error) that is expected to be most precise. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

*While specific anatomical hierarchies associated with PP are not detailed here, several suggestions for its neural instantiation have been proposed. See Bastos et al., 2012; Kanai, Komura, Shipp, & Friston, 2015; Shipp et al., 2013; Rauss & Pourtois, 2013.

am indeed drinking the water (Adams et al., 2013; FitzGerald, Schwartenbeck, Moutoussis, Dolan, & Friston, 2015; Friston et al., 2016; Pezzulo, Rigoli, & Friston, 2015; Schwartenbeck et al., 2015; Shipp, Adams, & Friston, 2013). Desires, on PP, are understood exclusively in terms of hypotheses – they are no different from perceptual hypotheses except in terms of their consequences.

Active and perceptual hypotheses are also intimately linked in minimizing free energy via the minimization of prediction error (Fig. 1). Perceptual inference provides information about the likely state of the world to active inference, which is used to initiate action or update proprioceptive predictions. Active inference in turn can be used to help minimize perceptual prediction error, such as when we move closer to the source of an ambiguous stimulus, or – as is our focus here – attend to a particular region or stimulus.

The idea of using BDT to explain action – including the sorts of dynamic action problems modelled in optimal control theory – is nothing new (Körding, 2007; Körding & Wolpert, 2006; Rescorla, 2016). Many reinforcement learning models are also Bayesian, in that prior information is expressed probabilistically and updated according to the rules of Bayesian inference (Ghavamzadeh, Mannor, Pineau, & Tamar, 2015). However, such models typically go beyond PP by including cost functions in order to allow agents to select the optimal action. Such cost functions are an additional theoretical posit not reducible to the hypotheses and prediction errors that form the basis of the PP's mental economy.

It is for this reason that PP theorists seek to demonstrate that active inference can provide the same results as optimal control theory and reinforcement learning without availing itself of such functions (Friston, Daunizeau, & Kiebel, 2009; Friston, Samothrakis, & Montague, 2012; Solway & Botvinick, 2012). The sorts of agent behaviours typically explained by reinforcement learning – those of maximizing utility or expected reward – are modelled using only active and perceptual inference. Prior expectations about occupying different states replace explicit representations of reward, and desired outcomes will be those that are more likely given the agent's generative model, replacing cost functions. Formally, the utility of an outcome is equated to its log prior probability. On this account, it is no longer assumed that agents act to maximize utility. They instead act to reduce prediction error (for a discussion of how this account reconceptualizes the role of dopamine in terms of precision expectations, see Colombo & Wright, 2017; Friston et al., 2009; Friston et al., 2012; Schwartenbeck, FitzGerald, Mathys, Dolan, & Friston, 2015). As we will see below, it is this attempt to do without cost functions that causes problems for the PP account of attention.

1.3. Attention as precision optimization

The PP account of attention provides a solution to the seemingly incongruous finding that in some cases prediction error is not attenuated even though the identity and location of the associated objects are highly predictable (Chaumon, Drouet, & Tallon-Baudry, 2008).The enhancement of prediction error signals in such circumstances is hypothesized to be the result of attention (Clark, 2013; Itti & Baldi, 2005), and is consistent with previous work demonstrating that attentional focus increases neural activation for relevant stimuli/locations in sensory cortices (Boynton, 2009; Brefczynski & DeYoe, 1999; Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; Gandhi, Heeger, & Boynton, 1999; Martínez et al., 1999; Reynolds & Heeger, 2009; Somers, Dale, Seiffert, & Tootell, 1999). A number of neuroscientists in recent years have adopted the PP theory in some capacity to elucidate mechanisms of attentional allocation (Chennu et al., 2013; Den Ouden, Kok, & De Lange, 2012; Feldman & Friston, 2010; Itti & Baldi, 2005; Jiang, Summerfield, & Egner, 2013; Kok, Rahnev, Jehee, Lau, & de Lange, 2012).

Descriptions of attention by PP theory, originally outlined by Feldman and Friston (2010) and defended by Hohwy (2012, 2013) and Clark (2013, 2015) state that attention is the optimization of the precision of prediction errors. In mathematical terms, precision is the inverse variance of a signal. Informally, precision can be thought of as a measure of the signal's reliability - how likely it is that the prediction error generated is the result of signal, as opposed to noise. However, precision will be context-dependent - in two different environmental contexts the prediction error generated by the same prediction may differ, because noise levels in the environment are apt to change. So just as we must learn to make perceptual predictions, so too must we make predictions about the likely precision of their corresponding prediction errors. Precision expectations are subjective estimates of how noisy or precise we expect the prediction error signal to be in a given context.

The development of precision expectations in vivo appears to be driven by learning statistical regularities about noise levels in the environment. For example, we learn that our vision (the prediction errors generated by our visual systems) is relatively imprecise in low lighting, or that our hearing is imprecise in environments with lots of background noise. Optimizing precision is the process of guiding hypothesis revision by directing processing resources towards the prediction errors with higher expected precisions – we attend to what is expected to be consistently most informative, and this information is used to preferentially revise our perceptual hypotheses. Such a practice allows us to avoid the potentially disastrous consequences of overfitting our hypotheses on the basis of noise-induced prediction errors.

The PP theory of attention is thus committed to the claim that high precision expectations are driving attention in all its instances. To date, the primary focus has been on employing the PP theory of attention to explain aspects of a canonical dichotomy that parses attention into ‘endogenous’ and ‘exogenous’ processes (Posner, 1980). Exogenous attention has been defined as an automatic orienting response to an environmental stimulus (Posner, 1980). Loud noises and bright flashes are paradigmatic examples of stimuli that ‘capture’ our attention, regardless of our desires or current activity. Within PP theory, this attentional capture is explained by a general standing expectation that large, abrupt prediction errors are typically highly precise (Clark, 2013, 2015;Feldman & Friston, 2010 ; Hohwy, 2012, 2013).

By contrast, the construct of endogenous attention involves the agent at least to some degree willfully directing resources towards some aspect of the environment, usually for a purpose or task (e.g. Hohwy, 2012). In empirical studies of the PP theory of attention, following suit with empirical studies of attention more generally, endogenous attention is often operationalized as task-relevance (Chennu et al., 2013; Jiang et al., 2013; Kok et al., 2012). Within PP theory, endogenous attention is explained by the learning of contextually sensitive regularities. For example, in the ‘Posner paradigm’ (Posner, 1980) – a cueing task designed to investigate how covert spatial attention can facilitate the detection of stimuli – one learns the regularity that arrows that point in a given direction will usually indicate that there is, or will be, a stimulus in that region. Over repeated exposure to this pairing, one comes to expect prediction errors generated for hypotheses indexed to a given spatial region to be highly precise when an arrow points to that region (Feldman & Friston, 2010; Hohwy, 2012, 2013). When the task is to detect a stimulus that will appear on the screen, one exploits this regularity to perform the task more efficiently.

Though less discussed than exogenous or endogenous attention, some PP theorists have also focused on ‘volitional’ or ‘voluntary’ attention – where attention is allocated on the basis of a conscious decision to do so (Hohwy, 2013, pp. 197–199). The construct of voluntary attention is meant to capture the phenomenon whereby we can attend to things simply because we decide to, without any external impetus. For example, if you were to decide on a whim that you want to attend to the fingernail on your pinkie finger, you could do so without any trouble. This form of attention is narrower than endogenous attention, though perhaps a subtype of it. While both endogenous and voluntary attention involve some degree of willful action on the part of the agent, endogenous attention needn't involve a conscious decision to attend.

Moreover, in the case of endogenous attention (construed as excluding volitional attention) the cue is at least partly external to the agent, whereas in the case of volitional attention the cue is wholly internal; the decision to attend to the cue is itself what serves to direct attention. The relevant learned regularity will be that our decisions to attend are highly correlated with the appearance of high precision targets (Hohwy, 2013).

2. The challenge from affect-biased attention

PP's focus on explaining endogenous and exogenous attention leaves out important attentional phenomena. Revisiting the case of the Doberman, it quickly becomes apparent that our increased attention to the yard is not accommodated by any of the explanations described above, yet this type of event can (and often ought to) capture our attention. It is not a case of exogenous attention as PP theory conceptualizes it, as there is nothing in the environment beyond the initial encounter that generates a large abrupt prediction error to capture our attention. Attention is allocated despite the lack of any such occurrence – we attend pre-emptively, despite the lack of any loud noises or sudden movement. Neither is it a case of endogenous attention as conceptualized by the PP theory. Given that it is unlikely the dog will be in the yard (recall its typical absence), the mere presence of the house or the fence does not serve to reliably indicate the presence of the dog. There is no learned statistical regularity here that can explain our attention. While the chance of the dog being present is higher in the yard than it is, say, in the supermarket or at the bank, it nevertheless remains quite low overall. The expected precision is thus also still relatively low.

Voluntary attention might seem at first to provide a solution – we desire not to be startled, so we make a conscious decision to attend to the fence in order to ensure there is no dog behind it. No external cue need be present in this case. However, there are two points that speak against this solution. First, it faces a similar problem as that of endogenous attention. We should not have a high precision expectation that our decision to attend will lead to the sighting of the dog, given the infrequency with which it is actually in the yard. Second, following the arguments of Ransom, Fazelpour, and Mole (2017), PP theory does not provide a coherent account of voluntary attention. Any increased precision in the sensory input will be merely a consequence of attending, rather than what is driving attention. This is because, while it is true that attending to something will, by a sort of 'self-fulfilling prophecy,' increase the precision of the prediction error associated with that hypothesis, it cannot be a precision expectation that is driving the decision to attend to that particular object. We will have equivalent precision expectations for all objects; no matter if we attend to this or that object, the precision will be enhanced in either case. So precision expectations cannot be what is responsible for driving voluntary attention (c.f. Clark, 2017). There is therefore no way to explain the selective orientation of attention in terms of precision expectations, given the current PP conceptualization of attention.

Part of the difficulty PP faces is that the classical separation of 'exogenous' and 'endogenous' attention is almost certainly too narrow a view of the operations of attention (Awh, Belopolsky, & Theeuwes, 2012; Todd & Manaligod, 2018). Specifically, affect-biased attention cannot be neatly classified into either category. Affect-biased attention is attention to stimuli that are affectively salient, i.e. stimuli that stand out because of their associations with reward or punishment (Markovic, Anderson, & Todd, 2014; Mather & Sutherland, 2011; Rolls, 2000; Todd, Cunningham, Anderson, & Thompson, 2012; Todd & Manaligod, 2018; Vuilleumier, 2015). It is not comfortably categorized as a form of exogenous attention because affectively salient objects, like other sources of attentional guidance driven by experience, can capture attention even when they are not physically salient (Anderson, 2013; Anderson, Laurent, & Yantis, 2011, 2012; Anderson & Yantis, 2013; Awh et al., 2012; Della Libera & Chelazzi, 2009; Hickey, Chelazzi, & Theeuwes, 2010; Libera & Chelazzi, 2006; Niu, Todd, & Anderson, 2012; Niu, Todd, Kyan, & Anderson, 2012; Shomstein & Johnson, 2013). Affectively salient objects also capture attention when they are not task relevant (Awh et al., 2012; Todd et al., 2012), preventing straightforward assimilation to PP's treatment of endogenous attention. It thus serves as an important area of investigation for those interested in the PP account of attention.

In addition, the Doberman scenario may be best described as involving affect-biased attention – we attend to the fence because it is affectively salient to us, regardless of whether we expect the dog to be there. In this case, it seems that a signal with a low chance of containing highly important information should be attended, though it is not expected to be precise on the PP theory. It also suggests the need to factor in the cost of false negatives and false positives, such that expected precisions can be weighted according to their cost. In an evolutionary context, there is often a significantly higher cost to false negatives compared to false positives – misidentifying a potentially dangerous object or situation as safe may have dire consequences, while the inverse is rather benign, though there are boundaries on this (Stephens, 2001).

The question is then whether factors such as the cost of getting it wrong can be accurately captured by precision expectations. We remain unconvinced that they can in all cases. As discussed in Section 1.2, PP theorists hold that cost functions can be eliminated in their framework by simply redescribing the utility of an outcome as its log prior probability. In cases where outcomes that have high utility also have high probabilities, then this redescription is not problematic. However, PP cannot accommodate cases where the two come apart. The example of affect-biased attention we have discussed here is one such case.

As defined in PP theory, precision expectations pertain to estimating not the cost of error, but the level of confidence that the sensory input itself is likely to be signal rather than noise. Estimating precision and the cost of getting things wrong diverge in many instances. Consider for example walking through the tall grass of the Ugandan savannah on a very windy day. The cost of mistaking the cause of a rustle in the grass as wind rather than a lion is quite high, yet the probability that the movement is caused by the wind is much higher than that it is caused by the lion. The rustling movement will therefore be expected to be relatively imprecise with respect to the hypothesis that it is caused by a lion. In cases such as these, precision expectations do not optimally drive attention in a manner required for survival (and presumably go against one's preference to remain in one piece).

The challenge for the PP theory that arises from the Doberman case, and from affect-biased attention more generally, is to accommodate these additional attentional influences in terms internal to the PP framework. Someone who endorses BDT might be puzzled at this point – why not just embrace cost functions so as to be able to readily accommodate the phenomenon? Of course, supporters of BDT must provide testable models demonstrating that cost functions do indeed address the issue. While a full discussion of how BDT might do so is beyond the scope of this paper given that there are many ways in which such models might be elaborated, and so evaluation will depend on these details, we suspect that a promising avenue will be to look to existing BDT models of sensorimotor control that posit two separate cost functions: error-based and reinforcement-based (Cashaback, McGregor, Mohatarem, & Gribble, 2017). Though attention is not perfectly understood as a form of sensorimotor control because some shifts of attention can be covert, the sensorimotor framework nevertheless provides general guidance for positing two different cost functions directing attentional processes. The first pertains to how precise we expect perceptual prediction errors to be, as hypothesized by PP theorists, where the cost function may be understood as minimizing prediction error (though, as observed by PP theorists, cost functions in this sense are theoretically unnecessary and can be fully assimilated as described above). The second pertains to how rewarding we expect a given perceptual outcome to be, where the cost function can be understood along traditional BDT lines as maximizing expected utility. This is where a

solution to the Doberman problem can be implemented, as expected utility is defined as:

$$E[\text{Utility}] \equiv \sum_{\substack{\text{possible} \\ \text{outcomes}}} p(\text{outcome} \mid \text{action}) U(\text{outcome})$$

Here *p(outcome|action)* is the probability of a given outcome given an action, which in this case, is the probability of the dog being present given the act of attending to the fence. *U(outcome)* is the utility or value that is associated with the outcome (or one might instead adopt prospect theory where value is instead assigned to prospective gains or losses (Tversky & Kahneman, 1979). As is evident from the equation, the expected utility of an action can still be relatively high when the probability of a given outcome is low, if the value of the outcome is high enough. We discuss some further work separating expected precision from expected utility, and highlight some directions for future research, in Section 5.

Regardless of the success of BDT at accommodating affect-biased attention, PP rejects cost functions because it purports to provide a unified and parsimonious account of mental functioning as minimizing free energy, not maximizing utility (see Section 1.1.4). To posit mechanisms beyond this framework would therefore invalidate what Colombo and Wright (2017) term the 'grand unified theory' that the main advocates of PP adhere to. It therefore remains to be shown whether and how affect-biased attention can be accommodated with only the internal tools at its disposition. In what follows we review the state of the art of empirical evidence for affect-biased attention, how PP has sought to accommodate affect so far using the theoretical tools it has at its disposal, and argue that these accounts do not resolve the challenge from affect-biased attention.

3. Affect and reward biases in attention

A large body of research demonstrates that affectively salient stimuli capture or guide attention (for a review see Pourtois, Schettino, & Vuilleumier, 2013). Affectively salient stimuli receive enhanced neural processing resources within sensory pathways (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012) including across various regions in the visual (Critchley et al., 2000; Damaraju, Huang, Barrett, & Pessoa, 2009; Morris et al., 1998; Padmala & Pessoa, 2008; Phan, Wager, Taylor, & Liberzon, 2002; Vuilleumier, Armony, Driver, & Dolan, 2001) and auditory cortices (Ethofer et al., 2006; Ethofer et al., 2012; Fecteau, Belin, Joanette, & Armony, 2007; Grandjean et al., 2005; Kryklywy, Macpherson, Greening, & Mitchell, 2013). Behaviorally, studies have demonstrated enhanced detection for emotional vs. neutral stimuli using paradigms such as visual search (Eastwood, Smilek, & Merikle, 2001; Kryklywy & Mitchell, 2014; Öhman, Flykt, & Esteves, 2001), spatial orienting (Armony & Dolan, 2002; Pourtois, Grandjean, Sander, & Vuilleumier, 2004) and attentional blink tasks (Anderson, 2005; De Martino, Kalisch, Rees, & Dolan, 2009; Lee, Todd, Gardhouse, Levine, & Anderson, 2013; McHugo, Olatunji, & Zald, 2013). Notably, such attentional biases do not need to be developed through extended life experience, but can be learned through conditioning. For example, appetitive conditioning studies demonstrate continued attentional priority allocated to stimulus features formerly associated with reward (Anderson et al., 2011; Chelazzi et al., 2014), an effect that endures across time regardless of continued reward pairing (e.g. Chelazzi et al., 2014).

Studies using EEG or MEG methods have shown enhanced event-related potentials (ERPs) for emotional stimuli at both early and late latencies following stimulus onset, including increases in C1 amplitude, an early visual cortical response reflecting low-level visual features, (Pourtois et al., 2004; Rauss, Schwartz, & Pourtois, 2011; Rossi & Pourtois, 2014; Stolarova, Keil, & Moratti, 2006; West, Anderson, Ferber, & Pratt, 2011) and increases in P1 amplitude, an index of extrastriate cortex activity typically enhanced for attended vs. unattended stimuli (Batty & Taylor, 2003; Pourtois et al., 2004; Pourtois, Dan, Grandjean, Sander, & Vuilleumier, 2005; Rotshtein et al., 2010). Enhanced processing in the visual cortex is also demonstrated by studies of steady-state visual evoked potentials (SSVEPs), where oscillatory neural activity frequency matched to that of an attended flickering stimulus is augmented for emotional imagery (Keil, Moratti, Sabatinelli, Bradley, & Lang, 2005; Müller, Andersen, & Keil, 2007; Wieser, McTeague, & Keil, 2012). Interestingly, while the enhanced ERPs for emotional stimuli reflect those of traditional endogenous and exogenous attention, fMRI and EEG research indicates that the sources of these effects are partially distinct from the regions typically noted to mediate endogenous attention and exogenous attention (fronto-temporal and temporoparietal respectively) (Corbetta & Shulman, 2002; Kastner & Ungerleider, 2001; Serences & Yantis, 2007). Notably, affect-biased attention recruits amygdala and midbrain circuitry not frequently implicated in other forms of attentional control (Todd & Manaligod, 2018). The amygdala has strong bidirectional connections with sensory areas (Amaral, Behniea, & Kelly, 2003; Catani, Jones, Donato, & Ffytche, 2003; Gschwind, Pourtois, Schwartz, Van De Ville, & Vuilleumier, 2012), and has been shown to play a role in guiding attention to rewarding (Peck, Lau, & Salzman, 2013) as well as punishing stimuli (Peck & Salzman, 2014; Todd & Manaligod, 2018; Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004) see also (Anderson & Phelps, 2001; Markovic et al., 2014; Pourtois et al., 2013). The locus coeruleus-norepinephrine (LC-NE) system in the midbrain biases attention to affectively salient stimuli by modulating visual cortical activity both directly and indirectly via the amygdala and ventromedial prefrontal cortex (vmPFC) (Aston-Jones & Cohen, 2005; Markovic et al., 2014; Mather, Clewett, Sakaki, & Harley, 2016).

While past research has disembedded affect-biased attention from both physical salience and task-based attention, it has not directly considered the role of precision expectations. It remains an open question how affect-biased attention interacts or competes with the sorts of precision expectations utilized during straightforward cases of endogenous and exogenous attention.

4. Preliminary attempts to incorporate affect into the PP framework

Here we review and evaluate three prominent attempts to provide a general account of our affective experience in the PP framework. We argue that none can accommodate affect-biased attention.

4.1. Embodied inference

Miller and Clark (2018) propose that precision expectations are extensively mediated by sub-cortical pathways, and they provide a review of the role of sub-cortical processing in modulating precision expectations and incorporating affect. For example, they follow (Pessoa, 2014) in viewing the medial pulvinar as amplifying weak but biologically valuable signals and so influencing behavior. Their picture is one in which action, affect, perception, and cognition are 'happily entangled' thanks to extensive thalamocortical loops, where "these subcortical loops help influence precision estimations in ways that reflect bodily states and unfolding actions, allowing value (to the organism) and affect (relating to interoceptive bodily states) to exert a continuous influence on high-level predictions" (Miller & Clark, 2018, p. 2572).

While Miller and Clark claim that such a story is consistent with PP (cf. Colombo & Wright, 2017), as far as we can see the evidence that they marshall in support of their thesis can actually be used to raise the same objections we have raised here. The fact that regions known to play a role in affective salience are also implicated in gain modulation does nothing to suggest that this modulation is best understood in terms of precision expectations. Precision expectations are second order hypotheses that provide a measure of our confidence in our first order predictions. Affect-biased attention continues to be problematic insofar as how significant an object is can differ from how confident we are that

the object is indeed present in our visual field. Precision expectations and affective salience can exert opposite influences on gain modulation.

What Miller and Clark need is a theoretical account of how affective predictions themselves generate prediction errors with variable precisions, but it is not easy to do this using only the tools provided by PP theory. Suppose that we predict that when a light appears on a screen and we press a button we will receive a reward – this is a first order prediction. We can also make predictions about the precision with which that relationship holds. For example, if pressing the button leads to reward 90% of the time, then we may form high precision expectations. If it only leads to reward 1% of the time then we may form low precision expectations. These low precision expectations would then lead to decreased attention to the task, according to PP. Accurate performance would likely slip, especially if the task is relatively difficult. But now suppose that the reward is $1000. This large benefit would plausibly cause us to increase attention to the task – the possibility of getting a large reward, however slight, is enough of a motivator to attend. Precision expectations regarding affectively salient stimuli will then suffer from the same original issue raised above: we attend in part on the basis of the cost of getting it wrong or right, not just on the basis of expected precisions. What we need is of course just the concept of expected utility or value. But then it is not expected precision alone that drives attention – costs and benefits also matter.

4.2. Interoceptive inference

Interoception is the ability to perceive internal bodily states and changes. A PP account of interoception, proposed by Seth (2013), suggests that we infer the causes of our interoceptive states in much the same way that we infer the causes of exteroceptive perceptual states (see also Seth, 2014; Seth & Friston, 2016; Seth, Suzuki, & Critchley, 2012). On this account, emotions are the product of top down predictions of interoceptive responses to external stimuli, which interact with bottom up interoceptive prediction errors. Interoceptive prediction errors can be minimized in one of two ways. First, analogously to perceptual inference, by revising predictions in light of prediction errors. For example, we may come to realize that we are feeling unexpectedly happy thanks to awareness of unpredicted interoceptive signals. Second, analogously to active inference, prediction errors can be minimized by changing the input so that it conforms to the hypothesis. For example, a drop in blood sugar levels may trigger the body to re-establish homeostasis by releasing additional glucose, or by prompting the agent to eat lunch.

This account thus makes room for affective or interoceptive predictions, which in some cases will be tied to their perceptual causes – we predict that the appearance of a loved one will cause interoceptive changes that amount to happiness, or, returning to our own example, that the Doberman's appearance will cause us to be startled and afraid. Precision expectations would then concern how likely the prediction error generated by actual interoceptive states is the product of noise. For example, one's heart rate might coincidentally rise for a reason unrelated to the presumed cause – such as it might if one ingested a large amount of caffeine – generating a prediction error that may be used to wrongly revise one's hypothesis that one is only mildly afraid of dogs (for an empirical investigation along these lines see Allen et al., 2016). Malfunctioning of the precision of interoceptive prediction errors has been hypothesized to play a role in chronic anxiety (Cornwell, Garrido, Overstreet, Pine, & Grillon, 2017; Paulus & Stein, 2006, 2010), as well as depersonalization and derealization disorders, where subjects feel persistently disembodied or as if the world around them is unreal (Seth et al., 2012); for other pathological cases hypothesized to involve the malfunctioning of the precision of prediction errors, see (Frith, 2012) (schizophrenia); (Adams et al., 2013) (schizophrenia & Parkinsonism); (Lawson, Rees, & Friston, 2014) (autism); (Seth & Friston, 2016) (autism, depression)).

What remains unclear is how such precision expectations might guide attention, and whether cases such as that of the Doberman can be accommodated. Given that our interoceptive hypotheses will be tied to their perceptual causes, it seems that in situations where some event is unlikely to occur, we should also not expect any interoceptive changes, nor should we expect high precision from any interoceptive changes that do occur. In cases where it is unlikely that the dog will appear there is no reason, on this account as it stands, to preemptively attend to the fence.

4.3. Emotional valence as the rate of change in free energy

In an account complementary to that of interoceptive inference discussed above, Joffily and Coricelli (2013) propose that emotional valence serves as a proxy for precision expectations. On their account, emotional valence is construed as the positive or negative character of emotion. Joffily & Coricelli propose that emotional valence – and a limited number of 'basic' emotions – be identified with the negative rate of change of free energy, where the minimization of free energy is roughly approximated by the minimization of prediction error under simplifying Gaussian assumptions.

On this account, emotional valence need not be explicitly represented by the organism. Rather, it is a consequence of how effectively free energy is minimized over time. While their main focus is on the valence component of emotions, they also provide an analysis of several emotions they term 'basic': happiness/unhappiness; hope/fear; relief/disappointment. For example, when free energy is decreasing (prediction error is being minimized) faster and faster over time, this produces the feeling of hope, insofar as the agent expects to be in a state of lower free energy in the near future. When free energy is decreasing more slowly over time the agent is said to be happy that her current state is one of lower free energy than the previous state. However, when free energy is increasing (accelerating) over time, this produces the feeling of fear, insofar as the agent expects to be in a state of higher free energy in the near future.

The connection with precision expectations (what Joffily & Coricelli term 'estimation uncertainty') is that instead of making direct estimations of the precision of one's hypotheses, the organism instead takes the increase in free energy as an indicator that it is overconfident in its hypotheses, and this results in an increased weighting of the prediction error. Likewise, when free energy is decreasing faster over time, then this will result in a decreased weighting of the prediction error. Intuitively, if our hypotheses are doing a poor job and producing lots of unresolved prediction error, then this suggests we ought to substantially revise them. On this account, valence serves as a sort of proxy for precision expectations.

Though the authors do not address attention, this account – when combined with the PP treatment of attention – predicts that negative valence will lead to enhanced attention, because it leads to enhanced gain on prediction error, and positive valence will lead to diminished attention. This account is almost certainly too simple insofar as we can and do attend to positively valenced events and objects (such as the $1000 reward example in Section 4.1). Additionally – putting the attentional gloss aside – it is not clear that the original account can accommodate the asymmetric weighting of gains and losses. Tversky & Kahneman's (1979) loss aversion principle states that agents assign a heavier weight to potential losses or penalties than to potential gains or rewards of equal size. In such cases, a loss of a given amount leads to a more extreme emotional response than an equal-sized gain (McGraw, Larsen, Kahneman, & Schkade, 2010). This is potentially problematic because there is no element of the account as it currently stands that predicts this asymmetry. Rather, the account predicts that free energy increasing or decreasing at the same rate from the same initial starting point should take on, respectively, negative and positive valences of the same magnitude, where magnitude might be thought to map on to another variable commonly thought to be partly constitutive of the emotions: arousal.

Readers may also worry that the account does not map onto the emotions identified by Joffily & Coricelli as 'basic'. For example, while fear is related to the prospect of visiting a state of increased free energy in the near future, there are plausibly cases where we fear things that are not associated with increases in free energy. There is no uncertainty as to whether a shot at the doctor's office will hurt, or whether we will actually receive the shot when we know it is coming. If an event is well-predicted then it seems that we ought to experience some form of positive emotion, given that free energy is decreasing. Joffilly & Coricelli suggest that this sort of issue may be resolved by appealing to different levels of the hierarchical generative model: while free energy may be decreasing or stationary at some levels, it may be increasing in others. This explanation works well for the example they discuss, that of a pedestrian getting hit by a bicycle. While the pedestrian expects to get hit in this particular circumstance – she sees the cyclist barreling towards her and realizes they are on a collision course – she does not expect to be injured while crossing the street, and it is this violation of expectations that produces the negative valence. Whether this account generalizes to the case of the shot seems questionable, as in this case one presumably also has an expectation that one will likely receive a shot when one goes to the doctor's office. Perhaps we might posit a more general (though somewhat ad hoc) standing expectation not to be injured, and posit that this is where the violation of expectations occurs.

Putting these problems aside, the larger issue relevant to our purposes is that the account does not provide a clear resolution to cases such as the Doberman problem. In addition to the basic emotions, Joffily & Coricelli also propose a distinction between fear and anxiety in terms of the presence or absence of the stimulus in the environment (2013, p.12). Fear – in addition to involving an increase in free energy as outlined above – pertains to perceiving threats that are present in the external environment. In the case of anxiety the environment is perceived as normal, yet the rate of free energy will increase at higher levels of the hierarchy due to abstract causes we associate with our sensations, producing anxiety.

Based on this analysis, the relevant emotion in the Doberman case is anxiety, given that the Doberman is not required to be present in order for one to direct one's attention to the fence. However, what is needed to complete the account is that the conditions for negative valence must themselves obtain in order to produce the anxiety in the first place; a level of the hierarchy at which prediction error is increasing. We do not expect the dog to be there, so this is a non-starter. Perhaps the mere presence of the fence – due to an association with the dog – causes a change in our interoceptive sensations such as heart rate and so on. Then issue is that our interoceptive sensations are well-predicted (we predict our heart rate will rise at the sight of the fence). Therefore, this should not lead to a failure to minimize prediction error – our sensations will change exactly as predicted. So there should be no acceleration in the rate of free energy, no anxiety, and no increased attention. The account provided will thus do better at explaining attention to unexpected sources of value or disvalue than expected sources of value or disvalue. When we walk by the house for the first time and the dog rushes the fence, then this unexpected change in our rate of prediction error minimization can cause both negative valence and increased attention. But the account does not accommodate the sorts of preemptive attentional behaviours discussed here.

5. Directions for future research

In summary, even though none of the approaches reviewed above can be straightforwardly applied to affect-biased attention or our problem case, they may nevertheless constitute promising directions of theoretical pursuit when understood in the context of a weakened version of the PP theory that does not aspire to the status of a grand unified theory of mental functioning.

While we remain skeptical of grand unified theories, we are hopeful that a weakened version of PP may emerge whereby prediction works alongside other mechanisms, perhaps drawing on resources in BDT. BDT offers a treatment of agential value in terms of utility functions, and guidelines for integrating utility functions with prior probabilities in order to optimize decision making (Körding, 2007). Though these tools have not yet been applied to the phenomenon of affect-biased attention, there is reason to suspect that they can adequately capture such attentional patterns so long as the subjective utilities are quantified in the right way. We recognize that this diminishes the attraction of the PP for some. Nevertheless, we think that this emerging account of the mind has much to offer. Bayesian perceptual modelling has proven remarkably successful at accounting for visual phenomena, and to some degree may also explain attentional phenomena. Though evidence for Bayesian perceptual modelling is not evidence for PP, and PP makes distinct theoretical commitments (see Section 1), there is nevertheless enough of a family resemblance such that empirical work using a PP paradigm can make a contribution to the BDT research program.

To date, empirical studies of attention and PP have typically held precision expectations static, with the attentional cue retaining a constant validity across trials (Chennu et al., 2013; Jiang et al., 2013; Kok et al., 2012). However, investigating the potential causal influence of expected precision on attentional behavior requires manipulating cue validity and examining the effect of this manipulation on attentional behavior. To truly accumulate empirical evidence for (or against) PP models of attention, studies must begin to explicitly alter such expectations, while keeping other factors such as the content of expectations constant. This will allow researchers to further investigate what is particular to PP compared to alternative BDT models, namely, the total dependence of different aspects of attentional selectivity on differential gradations of expected precisions.

Wyart, Nobre, and Summerfield (2012) have developed a paradigm that appears promising for manipulating each variable independently (see also Vossel, Geng, & Fink, 2014). Here, the main task of the experiment was to report whether a Gabor patch was present or absent in one of two presented locations, with the probe only appearing post-stimulus, the specific location indicated on a trial by trial basis. By providing cues about the block-level, location specific probability of the Gabor patch presence, and the trial level probability of the location to be probed, they were able to delineate the influences of signal probability and task relevance respectfully. Critically, this resulted in patterns of visual sensitivity that could differentiate between the impact of precision expectation and prediction error. This approach represents an exceptionally promising avenue forward to help isolate affect-biased attention as well.

In the case of affect-biased attention, then, assessing the explanatory adequacy of the PP theory of attention requires separating the effects of the associated reward on stimulus processing from those of expected precision. This could be achieved by distinct manipulations on the value of the reward associated with an object, on the one hand, and the validity of the cue indicating the presence of that object on the other. Manipulating these two variables independently should help us gain further understanding of how the two interact. For example, van Steenbergen et al. (2017) show that, at least with respect to certain classes of motor tasks, increasing reward increases the precision of perceptual representations in the task-relevant areas (extrastriate body area and fusiform face area). This would seem to suggest that reward serves as an independent modulator of gain, as we have argued above: precision expectations and affective salience are dissociable. However, the interplay between the two is far from clear, and a fascinating opportunity to understand exactly how the PP theory must be supplemented in order to achieve explanatory adequacy.

6. Conclusion

In this paper we have argued that affect-biased attention cannot be assimiliated to the PP theory of attention without significantly weakening that theory, perhaps by drawing on tools from BDT such as cost

functions. Affect-biased attention is not straightforwardly explained by the PP treatment of exogenous or endogenous attention, and it provides cases where precision expectations will be low but attention nevertheless ought to be directed to an object because of potential rewards or punishments. This suggests that in order to accommodate affect, PP theory must relinquish its claim that it provides a complete explanation of brain functioning.

We have reviewed three prominent attempts to account for affect within the PP framework. First, Miller and Clark (2018) highlight the role of the medial pulvinar in modulating the gain on weak signals that pertain to affectively significant stimuli. We argue that this is consistent with our claim that such gain modulation is not best understood in terms of precision expectations, and does nothing to resolve the challenge from affect-biased attention. Second, Seth (2013) proposes that we infer the causes of our internal sensations in much the same way that we infer the causes of our exteroceptive sensations, and that interoceptive inference can be identified with emotion. We argue that since our interoceptive hypotheses will be tied to their perceptual causes, cases of affectively salient yet unlikely events should not produce interoceptive changes, nor should we expect high precision from any interoceptive changes that do occur. Third, Joffily and Coricelli (2013) construe emotional valence as the negative rate of change of free energy, and thus a determinant of precision expectations. We argue that while the account can explain attention to unexpected sources of value or disvalue, it cannot explain our attention to expected sources of value or disvalue (that is, to affectively salient objects).

Finally, we discuss directions for future research on affect-biased attention and PP. Understanding how precision expectations and affective salience interact is critical to assessing the explanatory reach of PP theory. While BDT can in principle answer the challenge we have issued to PP here, the field would do well to provide some actual models of the neglected phenomenon.

CRediT authorship contribution statement

M.J. Rransom:Conceptualization, Writing - original draft, Writing - review & editing, Funding acquisition.**S. Fazelpour**:Conceptualization, Writing - original draft, Writing - review & editing, Funding acquisition.**J. Markovic**:Writing - original draft, Writing - review & editing, Funding acquisition.**J. Kryklywy**:Writing - review & editing.**E.T. Thompson**:Writing - review & editing.**R.M. Todd**:Writing - review & editing.

Declaration of competing interest

None.

Acknowledgment

We would like to thank Marisa Carrasco, Felipe De Brigard, Tobias Egner, Karl Friston, Walter Sinnott-Armstrong, and the attendees of the 2016 and 2017 Summer Seminar in Neuroscience and Philosophy (SSNAP) at Duke University for providing valuable feedback on our SSNAP-funded project on predictive processing and affect-biased attention. This paper is one of the outputs of our project.

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function, 218*(3), 611–643.

Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology, 46*, 219–227. https://doi.org/10.1016/j.conb.2017.08.010.

Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *eLife, 5*, Article e18103. https://doi.org/10.7554/eLife.18103.

Amaral, D., Behniea, H., & Kelly, J. (2003). Topographic organization of projections from the amygdala to the visual cortex in the macaque monkey. *Neuroscience, 118*(4), 1099–1120.

Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *Journal of Experimental Psychology: General, 134*(2), 258.

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature, 411*(6835), 305–309.

Anderson, B. A. (2013). A value-driven mechanism of attentional selection. *Journal of Vision, 13*(3), 1–16.

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences, 108*(25), 10367–10371.

Anderson, B. A., Laurent, P. A., & Yantis, S. (2012). Generalization of value-based attentional priority. *Visual Cognition, 20*(6), 647–658.

Anderson, B. A., & Yantis, S. (2013). Persistence of value-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 39*(1), 6.

Armony, J. L., & Dolan, R. J. (2002). Modulation of spatial attention by fear-conditioned stimuli: An event-related fMRI study. *Neuropsychologia, 40*(7), 817–826.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403–450.

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences, 16*(8), 437–443. https://doi.org/10.1016/j.tics.2012.06.010.

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research, 17*(3), 613–620.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76*(4), 695–711.

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology, 83*(2), 120–131.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association, 112*(518), 859–877.

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology, 76*, 198–211.

Boynton, G. M. (2009). A framework for describing the effects of attention on visual responses. *Vision Research, 49*(10), 1129–1143.

Brainard, D., & Gazzaniga, M. (2009). Bayesian approaches to color vision. *The Visual Neurosciences, 4*.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the'spotlight'of visual attention. *Nature Neuroscience, 2*(4), 370.

Cashaback, J. G. A., McGregor, H. R., Mohatarem, A., & Gribble, P. L. (2017). Dissociating error-based and reinforcement-based loss functions during sensorimotor learning. *PLoS Computational Biology, 13*(7), Article e1005623. https://doi.org/10.1371/journal.pcbi.1005623.

Catani, M., Jones, D. K., Donato, R., & Ffytche, D. H. (2003). Occipito-temporal connections in the human brain. *Brain, 126*(9), 2093–2107.

Chaumon, M., Drouet, V., & Tallon-Baudry, C. (2008). Unconscious associative memory affects visual processing before 100 ms. *Journal of Vision, 8*(3), 1–10.

Chelazzi, L., Eštočinová, J., Calletti, R., Gerfo, E. L., Sani, I., Della Libera, C., & Santandrea, E. (2014). Altering spatial priority maps via reward-based learning. *Journal of Neuroscience, 34*(25), 8594–8604.

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibánez, A., ... Bekinschtein, T. A. (2013). Expectation and attention in hierarchical auditory prediction. *Journal of Neuroscience, 33*(27), 11194–11205.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind.* Oxford University Press.

Clark, A. (2017). Predictions, precision, and agentive attention. *Consciousness and Cognition, 56*, 115–119. https://doi.org/10.1016/j.concog.2017.06.013.

Colombo, M., & Wright, C. (2017). Explanatory pluralism: An unrewarding prediction error for free energy theorists. *Brain and Cognition, 112*, 3–12.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science, 248*(4962), 1556–1559.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience, 3*(3), 201–215. https://doi.org/10.1038/nrn755.

Cornwell, B. R., Garrido, M. I., Overstreet, C., Pine, D. S., & Grillon, C. (2017). The unpredictive brain under threat: A neurocomputational account of anxious hypervigilance. *Biological Psychiatry, 82*(6), 447–454.

Critchley, H., Daly, E., Phillips, M., Brammer, M., Bullmore, E., Williams, S., ... Murphy, D. (2000). Explicit and implicit neural mechanisms for processing of social information from facial expressions: A functional magnetic resonance imaging study. *Human Brain Mapping, 9*(2), 93–105.

Damaraju, E., Huang, Y.-M., Barrett, L. F., & Pessoa, L. (2009). Affective learning enhances activity and functional connectivity in early visual cortex. *Neuropsychologia, 47*(12), 2480–2487.

De Martino, B., Kalisch, R., Rees, G., & Dolan, R. J. (2009). Enhanced processing of threat stimuli under limited attentional resources. *Cerebral Cortex, 19*(1), 127–133.

Della Libera, C., & Chelazzi, L. (2009). Learning to attend and to ignore is a matter of gains and losses. *Psychological Science, 20*(6), 778–784. https://doi.org/10.1111/j.1467-9280.2009.02360.x.

Den Ouden, H. E., Kok, P., & De Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology, 3*, 548.

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception &*

Psychophysics, 63(6), 1004–1013.
Elias, P. (1955). Predictive coding–I. *IRE Transactions on Information Theory, 1*(1), 16–24.
Ernst, M. O. (2010). Eye movements: Illusions in slow motion. *Current Biology, 20*(8), R357–R359.
Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., ... Wildgruber, D. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *Neuroreport, 17*(3), 249–253.
Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex, 22*(1), 191–200.
Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage, 36*(2), 480–487.
Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience, 4*, 215.
FitzGerald, T. H., Schwartenbeck, P., Moutoussis, M., Dolan, R. J., & Friston, K. (2015). Active inference, evidence accumulation, and the urn task. *Neural Computation, 27*(2), 306–328.
Friston, K. J. (2008). Hierarchical models in the brain. *PLoS Computational Biology, 4*(11), Article e1000211.
Friston, K. J. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences, 13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005.
Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS One, 4*(7), Article e6421. https://doi.org/10.1371/journal.pone.0006421.
Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics, 102*(3), 227–260. https://doi.org/10.1007/s00422-010-0364-z.
Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O'Doherty, J., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews, 68*, 862–879. https://doi.org/10.1016/j.neubiorev.2016.06.022.
Friston, K. J., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology, Paris, 100*(1–3), 70–87.
Friston, K. J., Samothrakis, S., & Montague, R. (2012). Active inference and agency: Optimal control without cost functions. *Biological Cybernetics, 106*(8), 523–541. https://doi.org/10.1007/s00422-012-0512-8.
Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., ... Bestmann, S. (2012). Dopamine, affordance and active inference. *PLoS Computational Biology, 8*(1), Article e1002327.
Frith, C. (2012). Explaining delusions of control: The comparator model 20 years on. *Consciousness and Cognition, 21*(1), 52–54.
Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences, 96*(6), 3314–3319. https://doi.org/10.1073/pnas.96.6.3314.
Gershman, S.J., 2019. What does the free energy principle tell us about the brain? arXiv preprint arXiv:1901.07945.
Ghavamzadeh, M., Mannor, S., Pineau, J., & Tamar, A. (2015). Bayesian reinforcement learning: A survey. *Foundations and Trends® in Machine Learning, 8*(5–6), 359–483.
Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8*(2), 145–146.
Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., & Vuilleumier, P. (2012). White-matter connectivity between face-responsive regions in the human brain. *Cerebral Cortex, 22*(7), 1564–1576.
Heeger, D. J. (2017). Theory of cortical function. *Proceedings of the National Academy of Sciences, 114*(8), 1773–1782.
Helmholtz, H.v. (2005). *Treatise on physiological optics.* Mineola: Dover.
Hickey, C., Chelazzi, L., & Theeuwes, J. (2010). Reward changes salience in human vision via the anterior cingulate. *The Journal of Neuroscience, 30*(33), 11096–11103. https://doi.org/10.1523/jneurosci.1026-10.2010.
Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology, 3*(96), https://doi.org/10.3389/fpsyg.2012.00096.
Hohwy, J. (2013). *The predictive mind.* Oxford: Oxford University Press.
Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. Cognition, 108(3), 687-701. doi: https://doi.org/10.1016/j.cognition.2008.05.010.
Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature, 436*(7047), 71.
Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(5), 580–593. https://doi.org/10.1002/wcs.142.
Itti, L., & Baldi, P. F. (2005). Bayesian surprise attracts human attention. Paper presented at the proceedings of the 18th international conference on neural information processing systems (NIPS'05).
Jiang, J., Summerfield, C., & Egner, T. (2013). Attention sharpens the distinction between expected and unexpected percepts in the visual brain. *Journal of Neuroscience, 33*(47), 18438–18447.
Joffily, M., & Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS Computational Biology, 9*(6), Article e1003094.
Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1668), 20140169.
Kastner, S., & Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia, 39*(12), 1263–1276.
Keil, A., Moratti, S., Sabatinelli, D., Bradley, M. M., & Lang, P. J. (2005). Additive effects of emotional content and spatial selective attention on electrocortical facilitation. *Cerebral Cortex, 15*(8), 1187–1197.
Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference: Cambridge University Press*.
Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & de Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cereb Cortex*, bhr310.
Körding, K. P. (2007). Decision theory: What "should" the nervous system do? *Science, 318*(5850), 606–610. https://doi.org/10.1126/science.1142998.
Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. Trends in Cognitive Sciences, 10(7), 319-326. doi: https://doi.org/10.1016/j.tics.2006.05.003.
Kryklywy, J. H., Macpherson, E. A., Greening, S. G., & Mitchell, D. G. (2013). Emotion modulates activity in the "what" but not "where" auditory processing pathway. *Neuroimage, 82*, 295–305.
Kryklywy, J. H., & Mitchell, D. G. (2014). Emotion modulates allocentric but not egocentric stimulus localization: Implications for dual visual systems perspectives. *Experimental Brain Research, 232*(12), 3719–3726.
Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience, 8*, 302.
Lee, D., Todd, R., Gardhouse, K., Levine, B., & Anderson, A. (2013). *Enhanced attentional capture in survivors of a single traumatic event. Paper presented at the Society for Neuroscience Annual Meeting.* CA, USA: San Diego.
Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A, 20*(7), 1434–1448. https://doi.org/10.1364/JOSAA.20.001434.
Libera, C. D., & Chelazzi, L. (2006). Visual selective attention and the effects of monetary rewards. *Psychological Science, 17*(3), 222–227. https://doi.org/10.1111/j.1467-9280.2006.01689.x.
Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences, 35*(3), 121–143.
MacKay, D. J. (1995). Free energy minimisation algorithm for decoding and cryptanalysis. *Electronics Letters, 31*(6), 446–447.
Mamassian, P., Landy, M., & Maloney, L. T. (2002). Bayesian modelling of visual perception. *Probabilistic Models of the Brain*, 13–36.
Markovic, J., Anderson, A. K., & Todd, R. M. (2014). Tuning to the significant: Neural and genetic processes underlying affective enhancement of visual perception and memory. *Behavioural Brain Research, 259*, 229–241. https://doi.org/10.1016/j.bbr.2013.11.018.
Martínez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., ... Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience, 2*, 364. https://doi.org/10.1038/7274.
Mather, M., Clewett, D., Sakaki, M., & Harley, C. W. (2016). Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behavioral and Brain Sciences, 39*.
Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science, 6*(2), 114–133. https://doi.org/10.1177/1745691611400234.
McGraw, A. P., Larsen, J. T., Kahneman, D., & Schkade, D. (2010). Comparing gains and losses. *Psychological Science, 21*(10), 1438–1445. https://doi.org/10.1177/0956797610381504.
McHugo, M., Olatunji, B., & Zald, D. (2013). The emotional attentional blink: What we know so far. *Frontiers in Human Neuroscience, 7*(151), https://doi.org/10.3389/fnhum.2013.00151.
Miller, M., & Clark, A. (2018). Happily entangled: Prediction, emotion, and the embodied mind. *Synthese, 195*(6), 2559–2575.
Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proceedings of the National Academy of Sciences, 108*(30), 12491–12496.
Morris, J. S., Friston, K. J., Büchel, C., Frith, C. D., Young, A. W., Calder, A. J., & Dolan, R. J. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain: A Journal of Neurology, 121*(1), 47–57.
Müller, M. M., Andersen, S. K., & Keil, A. (2007). Time course of competition for visual processing resources between emotional pictures and foreground task. *Cerebral Cortex, 18*(8), 1892–1899.
Niu, Y., Todd, R., & Anderson, A. (2012). Affective salience can reverse the effects of stimulus-driven salience on eye movements in complex scenes. *Frontiers in Psychology, 3*(336), https://doi.org/10.3389/fpsyg.2012.00336.
Niu, Y., Todd, R. M., Kyan, M., & Anderson, A. K. (2012). Visual and emotional salience influence eye movements. *ACM Transactions on Applied Perception, 9*(3), 1–18. https://doi.org/10.1145/2325722.2325726.
Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General, 130*(3), 466.
Padmala, S., & Pessoa, L. (2008). Affective learning enhances visual detection and responses in primary visual cortex. *Journal of Neuroscience, 28*(24), 6202–6210.
Paulus, M. P., & Stein, M. B. (2006). An insular view of anxiety. *Biological Psychiatry, 60*(4), 383–387.
Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Structure and Function, 214*(5–6), 451–463.
Peck, C. J., Lau, B., & Salzman, C. D. (2013). The primate amygdala combines information about space and value. *Nature Neuroscience, 16*(3), 340–348. https://doi.org/10.1038/nn.3328.
Peck, C. J., & Salzman, C. D. (2014). Amygdala neural activity reflects spatial attention towards stimuli promising reward or threatening punishment. *eLife, 3*, Article e04478.
Pessoa, L. (2014). Understanding brain networks and brain organization. *Physics of Life Reviews, 11*(3), 400–435.
Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. Progress in Neurobiology, 134, 17-35. doi: https://doi.org/10.1016/j.pneurobio.2015.09.001.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage, 16*(2), 331–348.
Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3–25.
Pourtois, G., Dan, E. S., Grandjean, D., Sander, D., & Vuilleumier, P. (2005). Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: Time course and topographic evoked-potentials mapping. *Human Brain Mapping, 26*(1), 65–79.
Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cerebral Cortex, 14*(6), 619–633. https://doi.org/10.1093/cercor/bhh023.
Pourtois, G., Schettino, A., & Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention: What is magic and what is not. *Biological Psychology, 92*(3), 492–512. https://doi.org/10.1016/j.biopsycho.2012.02.007.
Ransom, M., Fazelpour, S., & Mole, C. (2017). Attention in the predictive mind. *Consciousness and Cognition, 47*, 99–112.
Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79.
Rauss, K., Schwartz, S., & Pourtois, G. (2011). Top-down effects on early visual processing in humans: A predictive coding framework. *Neuroscience & Biobehavioral Reviews, 35*(5), 1237–1253.
Rauss, K., & Pourtois, G. (2013). What is bottom-up and what is top-down in predictive coding? *Frontiers in psychology, 4*, 276.
Rescorla, M. (2016). Bayesian sensorimotor psychology. *Mind & Language, 31*(1), 3–36.
Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. Neuron, 61(2), 168-185. doi: https://doi.org/10.1016/j.neuron.2009.01.002.
Rohe, T., & Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biology, 13*(2), Article e1002073.
Rolls, E. T. (2000). On the brain and emotion. *Behavioral and Brain Sciences, 23*(2), 219–228. https://doi.org/10.1017/S0140525X00512424.
Rossi, V., & Pourtois, G. (2014). Electrical neuroimaging reveals content-specific effects of threat in primary visual cortex and fronto-parietal attentional networks. *Neuroimage, 98*, 11–22.
Rotshtein, P., Richardson, M. P., Winston, J. S., Kiebel, S. J., Vuilleumier, P., Eimer, M., ... Dolan, R. J. (2010). Amygdala damage affects event-related potentials for fearful faces at specific time windows. *Human Brain Mapping, 31*(7), 1089–1105.
Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., & Friston, K. (2015). The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cerebral Cortex, 25*(10), 3434–3445.
Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., Dolan, R., Kronbichler, M., & Friston, K. (2015). Evidence for surprise minimization over value maximization in choice behavior. *Scientific Reports, 5*, 16575. https://doi.org/10.1038/srep16575.
Serences, J. T., & Yantis, S. (2007). Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. *Cerebral Cortex, 17*(2), 284–293.
Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences, 17*(11), 565–573.
Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience, 5*(2), 97–118.
Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 371*(1708), 20160007.
Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology, 2*, 395.
Shipp, S., Adams, R. A., & Friston, K. J. (2013). Reflections on agranular architecture: Predictive coding in the motor cortex. Trends in Neurosciences, 36(12), 706-716. doi: https://doi.org/10.1016/j.tins.2013.09.004.
Shomstein, S., & Johnson, J. (2013). Shaping attention with reward: Effects of reward on space- and object-based selection. *Psychological Science, 24*(12), 2369–2378. https://doi.org/10.1177/0956797613490743.
Solway, A., & Botvinick, M. M. (2012). Goal-directed decision making as probabilistic inference: A computational framework and potential neural correlates. *Psychological Review, 119*(1), 120.
Somers, D. C., Dale, A. M., Seiffert, A. E., & Tootell, R. B. H. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proceedings of the National Academy of Sciences, 96*(4), 1663–1668. https://doi.org/10.1073/pnas.96.4.1663.
Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: A predictive coding view. *Frontiers in Human Neuroscience, 8*(666), https://doi.org/10.3389/fnhum.2014.00666.
Stephens, C. L. (2001). When is it selectively advantageous to have true beliefs? Sandwiching the better safe than sorry argument. *Philosophical Studies, 105*(2), 161–189. https://doi.org/10.1023/a:1010358100423.
Stolarova, M., Keil, A., & Moratti, S. (2006). Modulation of the C1 visual event-related component by conditioned stimuli: Evidence for sensory plasticity in early affective perception. *Cerebral Cortex, 16*(6), 876–887.
Stone, J. V. (2011). Footprints sticking out of the sand. Part 2: Children's Bayesian priors for shape and lighting direction. *Perception, 40*(2), 175–190.
Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron, 59*(2), 336–347.
Todd, R. M., Cunningham, W. A., Anderson, A. K., & Thompson, E. (2012). Affect-biased attention as emotion regulation. *Trends in Cognitive Sciences, 16*(7), 365–372. https://doi.org/10.1016/j.tics.2012.06.003.
Todd, R. M., & Manaligod, M. G. M. (2018). Implicit guidance of attention: The priority state space framework. Cortex, 102, 121-138. doi: https://doi.org/10.1016/j.cortex.2017.08.001.
Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *Journal of Neuroscience, 32*(39), 13389–13395.
Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. *Journal of Neuroscience, 31*(25), 9118–9123.
Tversky, A., & Kahneman, D. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291.
van Steenbergen, H., Warren, C. M., Kühn, S., de Wit, S., Wiers, R. W., & Hommel, B. (2017). Representational precision in visual cortex reveals outcome encoding and reward modulation during action preparation. *NeuroImage, 157*, 415–428.
Vossel, S., Geng, J. J., & Fink, G. R. (2014). Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles. *The Neuroscientist, 20*(2), 150–159.
Vuilleumier, P. (2015). Affective and motivational control of vision. *Current Opinion in Neurology, 28*(1), 29–35.
Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron, 30*(3), 829–841.
Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience, 7*(11), 1271.
Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience, 32*(11), 3665–3678.
Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6), 598.
West, G. L., Anderson, A. A., Ferber, S., & Pratt, J. (2011). Electrophysiological evidence for biased competition in V1 for fear expressions. *Journal of Cognitive Neuroscience, 23*(11), 3410–3418.
Wieser, M. J., McTeague, L. M., & Keil, A. (2012). Competition effects of threatening faces in social anxiety. *Emotion, 12*(5), 1050–1060. https://doi.org/10.1037/a0027069.
Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences, 109*(9), 3593–3598. https://doi.org/10.1073/pnas.1120118109.