Consciousness and Cognition 47 (2017) 99–112

Contents lists available at ScienceDirect

Consciousness and Cognition

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

Attention in the predictive mind

Madeleine Ransom, Sina Fazelpour, Christopher Mole *

Department of Philosophy, University of British Columbia, Vancouver, Canada

ARTICLE INFO

Article history:
Received 30 November 2015
Revised 28 May 2016
Accepted 13 June 2016
Available online 4 July 2016

Keywords:
Prediction-error coding
Attention
Voluntary attention
Philosophy of perception
Hohwy

ABSTRACT

It has recently become popular to suggest that cognition can be explained as a process of Bayesian prediction error minimization. Some advocates of this view propose that attention should be understood as the optimization of expected precisions in the prediction-error signal (Clark, 2013, 2016; Feldman & Friston, 2010; Hohwy, 2012, 2013). This proposal successfully accounts for several attention-related phenomena. We claim that it cannot account for all of them, since there are certain forms of *voluntary* attention that it cannot accommodate. We therefore suggest that, although the theory of Bayesian prediction error minimization introduces some powerful tools for the explanation of mental phenomena, its advocates have been wrong to claim that Bayesian prediction error minimization is 'all the brain ever does'.

1. Prediction-error coding

An enormous amount of research has been devoted, over the last sixty years, to the discovery of efficient methods for gathering and disseminating information. The concepts that have emerged from this research have often been applied in theories of the mind. One of them enables us to draw a distinction between two quite different strategies for information gathering.

To see that distinction, suppose that I want to gather information about *x* from those of my sources that are in touch with *x*'s activities. There are two broad strategies that I might employ.[1] The first is to have my sources tell me some of the things that they know about *x*. The second is for me to tell those sources what I already think is going on with *x*, and to have them tell me some of the ways in which I am mistaken. If my sources and I are using this second strategy then any message that I receive from them will tell me only about ways in which my prior predictions were erroneous. This strategy is therefore called 'prediction error coding'.

In some contexts the first strategy will be the most efficient, in the sense that it will require the fewest bits of information to be transmitted. In other contexts the more efficient strategy will be the second. Since the systems that employ this strategy do not need their sources to transmit any parts of the gathered information that have already been correctly predicted, the second strategy will be especially efficient when our prior expectations are largely accurate. Such efficiencies can be a significant advantage when the transmission of information incurs a cost.

* Corresponding author at: UBC Department of Philosophy, 1866 Main Mall, E370, Vancouver, BC V6T 1Z1, Canada.
E-mail address: cmole@mail.ubc.ca (C. Mole).

[1] In this section we consider information gathering quite generally. In the section that follows we move to considering the brain as an information-gathering system. 'Guesses' will then be understood as corresponding to the top-down hypotheses, generated by an internal model; the information collected by sources will be understood as corresponding to the signals received in various sensory modalities. The mismatch between the two corresponds to the prediction error, which is passed up the hierarchy and used to revise the top-down predictions. Before considering its merits as an explanation of cognition, however, it is worth considering prediction-error minimization as a general strategy for information handling.

http://dx.doi.org/10.1016/j.concog.2016.06.011
1053-8100/

Upon receiving a signal telling me about the ways in which my prior predictions were erroneous, I can then update those predictions, so that the information from this incoming signal is incorporated into them. If I perform this updating in a way that takes account of my prior confidence in those predictions, as well as their successes or failures in accounting for the newly received information – and if I avoid violating the probability calculus while doing so – then my prediction-updating process will take the form of a Bayesian inference. The information-gathering process can then be repeated, with my sources again providing information about any new or remaining errors in my now-updated hypotheses. This reiterated process of updating and re-checking is referred to in the world of informatics as 'prediction error minimization' (e.g., Bishop, 2006; Moore & Weiss, 1979). In the world of criminal and military intelligence it is central to what is known as 'the intelligence cycle' (e.g., Richards, 2010, p. 10). There are various ways in which the magnitude of prediction errors can be calculated, and various ways in which the process of minimizing those errors can be implemented.[2] All of these employ variations on the same Bayesian logic.

If my information-gathering operation is of sufficient complexity then my hypotheses about what might be going on with *x* may be quite abstract, and may be concerned with a fairly long time window, whereas the information that is available to my information-gathering operatives in the field will be more concrete, and more short-term. This difference as to the levels at which we are operating has the potential to be problematic, since the information provided by my operatives may be at the wrong level of specificity for the testing of my high-level hypotheses. I might believe that *x*'s business is in some way expanding, but my information-gathering man on the street might not be in a position to assess that expansion directly. He might only be able to provide information about some more fine-grained hypotheses, concerning such concrete and short-term matters as the number of delivery vehicles leaving *x*'s compound on one particular evening. My high-level hypotheses may not say very much about these specific matters.

To prevent this difference of level from becoming problematic, the information-gathering networks in which prediction error coding is employed can be arranged in a hierarchical structure. Rather than asking the operative-on-the-street to correct my high level hypotheses about how *x*'s business is doing, I can instead get those hypotheses corrected by some intermediary analyst, and this intermediary can be the one who generates the lower level hypotheses that are to be tested by my man on the street. The number of these intermediary stages might sometimes need to be large (and there might also be strategic reasons why the use of several intermediaries is convenient), but the same logic of Bayesian hypothesis testing and prediction error coding can be used at each level of the hierarchy that results from the introduction of these intermediaries. Processes dealing with more abstract hypotheses – predicting changes over a relatively long spatiotemporal scale – are (ipso facto) counted as occupying a high level in such a hierarchy. Processes dealing with less abstract hypotheses – which will typically predict changes over a relatively short spatiotemporal scale – are (ipso facto) counted as occupying a low level in these hierarchies (Friston, 2008; Harrison, Bestmann, Rosa, Penny, & Green, 2011; Hohwy, 2012, 2016; Kiebel, Daunizeau, & Friston, 2008).[3] Each level sends its hypotheses to the level below, from which it in turn receives information about the things that these hypotheses have failed to predict.

1.1. Prediction errors in the brain

In systems where the transmission and storage of information incurs a cost, there will often be benefits to employing a strategy of hierarchically-organized Bayesian inference, through which prediction error is minimized. Given the metabolic costs of neural activity (Barlow, 1972), the brain can be thought of as one such system. A number of philosophers and psychologists have recently been exploring the idea that it is one of the places in which hierarchical prediction-error minimization is employed (e.g., Dayan, Hinton, Neal, & Zemel, 1995; Doya, 2007; Friston, 2009; Huang & Rao, 2011; Lee & Mumford, 2003; Rao & Ballard, 1999; Rao, Olshausen, & Lewicki, 2002; Stefanics, Kremlácek, & Czigler, 2014). The philosophers who have gone furthest in developing this approach are Jakob Hohwy and Andy Clark. It is on Hohwy's treatment that the present essay focuses, especially as given in his 2013 book, *The Predictive Mind*. Page references in what follows are to that book, unless otherwise noted.[4]

To say that the brain depends on prediction-error coding in its information-gathering operations is to say that perceptual systems do not provide the rest of the brain with *positive* information about how the world is. Their only job is to remove falsities from the brain's most-recently updated picture of that world, by providing information about the ways in which this

[2] The success or failure of predictions in light of what the sources tell me can be formalized in terms of likelihood functions or distance functions, such as the Kullback-Leibler divergence (Bishop, 2006; Gelman, Carlin, Stern, & Rubin, 2014). In an ideal situation, the prediction error would equal zero: given my guess, the response will be completely anticipated, and so there will be no reason to change the guess.

[3] See, however, Vance (2015) for an objection against this commitment of predictive error minimization theory: higher levels of the hierarchy needn't always correspond to hypotheses concerned with larger timescales or more abstract matters.

[4] We focus on Hohwy's account partly for the sake of expository convenience. Some versions of our argument may be applicable to other theories that are committed to the claim that attention is to be identified with the optimization of precision (Clark, 2013, 2016; Feldman & Friston, 2010; Friston, 2009, 2010), but our argument, as we develop it here, is directed against Hohwy's particular Bayesian treatment of attention, and not against 'Bayesian brain' theories *tout court*. 'Bayesianism' names a broad approach, within the scope of which various explanatory moves can be made. So broad a framework is best adjudicated via an adjudication of the particular theories that operate within it. The present paper makes one local contribution to that broader adjudicatory project.

picture is mistaken. Perception is therefore said by Hohwy to be *indirect*, "in the sense that what you experience *now* is given in your top-down prediction of your ongoing sensory input" (p. 48). That 'top-down prediction' is itself understood to have been generated through an empirical Bayesian process, in which a hierarchy of hypotheses were generated and tested (see Carlin & Louis, 2000), and in which prior probabilities were derived from a hierarchically organized Bayesian inference (Friston, 2005). The hypotheses that are formulated in such a system can be characterized with probability density functions in which probabilities are assigned to the variables gauged at lower levels in the hierarchy, and ultimately to the various possible causes of our sensations. The processing of the sensory signals against which these hypotheses are eventually tested is processing of the same Bayesian sort. Prediction-error coding and Bayesian hypothesis-testing are therefore be taken to operate at every level of the brain's processing hierarchy, with the result that Bayesian prediction error minimization "gives the organizing principle for brain function as such" (p. 101).

As these last remarks indicate, the prediction-error minimization theory has ambitions to completeness, and purports to provide a unified explanation for what would otherwise be a disparate range of psychological phenomena, including phenomena of perception and of action. These ambitions are a source of controversy (Colombo & Wright, 2015), but they are also claimed as one of the theory's main advantages. Andy Clark writes that:

> Such approaches, originally developed in the domain of perception, have been extended [. . .] to encompass action, and to offer an attractive, unifying perspective on the brain's capacities for learning, inference, and the control of plasticity. Perception and action, if these unifying models are correct, are intimately related and work together to reduce prediction error by sculpting and selecting sensory inputs.
>
> [Clark, 2013, p. 3]

Among the several things for which this theory aims to give a unified explanation are the phenomena of attention. This is made explicit at the beginning of *The Predictive Mind*, when Hohwy tells us that 'Perception, action *and attention* are but three different ways of doing the very same thing.' (p. 2, emphasis added). In Chapter Nine of that book he goes on to claim that:

> The prediction error framework is able to encompass many of the central findings on attention and to provide a unifying framework for them within the broader church of prediction error minimization. This allows us to see attention in a new light and to provide alternative conceptualizations of its functional role in our overall mental economy.
>
> [p. 191]

The present essay is an attempt to gauge the success of this 'alternative conceptualization'. We claim that its success is only partial. Hohwy's is a theory in which "A single type of mechanism, reiterated throughout the brain, manages everything" (p. 2). We suggest that this mechanism struggles to account for the full range of ways in which attention comes to be allocated. This does not imply that the prediction error minimization hypothesis is wrong. Much of what we say below is an advertisement of that hypothesis' strengths. Our negative claim is only that there are limits to the range of attention-related phenomena that can be explained by adopting it.

One might see this as a partial vindication of the theory that is based on prediction error minimization, but that theory's advocates will take a partial vindication to be no vindication at all. That is because they take their theory's claim to completeness to be the source of its philosophical interest. Hohwy writes that "Many would agree with the general idea that predictions play a role in perception but few would agree that prediction error minimization is all that the brain ever does and that action and attention is nothing but such minimization" (p. 7). Insofar as the agenda of the present article is a negative one, it is this 'nothing but' with which we are taking issue. There are some phenomena of attention that the prediction error theory handles nicely, but there are others – especially those involving the voluntary control of attention – with which it struggles.

A reviewer for this journal reminds us that a theory might claim to be complete, as an account of some organ's function, without needing to claim that all features of that organ's behaviour are explained by it: Our theory of the heart's function need only be compatible with our explanation of its behaviour in response to shock. That explanation need not be given by this theory alone. Analogously, a theory saying that the prediction error minimization is the only function of the brain need not provide an explanation for everything the brain does. It need only be compatible with whatever explanations are given. This point is well taken. The prediction error minimization theory could claim to have offered a complete theory of the brain's function, without claiming to have offered a complete explanation of its behaviour. The explanations that remain to be given should, however, be compatible with that theory of function, and – for reasons that the final section of this paper attempts to make clear – the still-to-be-completed explanatory work may, in the present case, require resources that are different from any that a Bayesian theory admits.

2. Surprisal and selection

We have said that systems employing prediction error coding owe some of their efficiency to the fact that the only information to be propagated up through such systems is 'surprising', in the sense that it is information about those aspects of the incoming signal that have not already been predicted by the system's prior hypotheses (see for example Barto, Mirolli, & Baldassarre, 2013; Friston, 2009). It is this that prevents resources from being taken up with the transmission of redundant information.

Some psychologists – following an earlier tradition (e.g., Moray, 1967) – are willing to apply the label of 'attention' to any feature of the brain's information handling that is implicated in the conservative allocation of its processing resources (e.g., Carrasco, 2011). These psychologists may be willing to say that this built-in redundancy-avoidance gives the systems that employ prediction-error coding a built-in mechanism of attention.[5] An idea along broadly these lines has been developed by Itti and Baldi (2006), who offer a model in which Bayesian prediction errors are used to explain the allocation of visual attention to those spatial regions concerning which our predictions are most erroneous.

Itti and Baldi's work gives an accurate model for many of the cases in which attention is allocated to spatially defined regions on the basis of external cues, but few advocates of the prediction-error framework would suppose that this approach could be generalized into a theory of all of the ways in which attention is allocated. That is because a built-in avoidance of redundancy is integral to the use of prediction-errors as a strategy for the encoding of information. It gets applied automatically, whenever that strategy is employed. It therefore gets applied uniformly. There is nothing here with which to explain those features of top-down attention that are involved in the *selective* direction of attention, first to some task-relevant things and then to others. And this selectivity is one of the things that any complete theory of attention would need to explain.

In order to bring such selectivity into view, it will be helpful to consider a paradigmatic example of it. The example on which we focus below is given by Ulric Neisser and Robert Becklen's now-classic 1975 study of 'selective looking'.

Neisser and Becklen presented their participants with two overlapping streams of visual information, both of which appeared in the same portion of their visual field. They did this by using a system of half-silvered mirrors, thereby creating an effect that is something like the effect of looking out through the window of a lit room at dusk, when the exterior and reflected interior worlds are both visible in the same part of space. Neisser and Becklen's schematic representation of their stimuli is reproduced in Fig. 1. It should be noted that the two films presented to their participants were quite different from each other. Both depicted the playing of a game, but one was filmed from up close and the other from further away. There was therefore little chance of perceiving their combination as a single, doubly-populated scene.

Neisser and Becklen found that people are quite capable of attending to either one of these scenes while ignoring the other; that we are quite capable of switching our attention between the scenes, provided that both are presented to both eyes; and that we are bad at attending to both scenes at the same time. Their experiment is a useful example for our present purposes because the 'selective looking' that it involves is clearly a form of selective *attention*, rather than being a case of priming, of perceptual-readiness, or of the mere orientation of one's sense organs. Some of this experiment's significant features are highlighted when Neisser and Becklen write that:

> Without any prior practice, it is easy to attend to one sequential episode and ignore another, even when both are presented in haphazard overlap. What defines one episode and distinguishes it from the other is not its distance or its clarity, or the sense organ involved, but its intrinsic properties and structure. [...] It is implausible to suppose that special "filters" or "gates," designed on the spot for this novel situation, block the irrelevant material from penetrating deeply into the processing system. The ordinary perceptual skills of following visually-given events, which develop in the first year of life (Bower, 1971), are simply applied to the attended episode and not to the other.
>
> [Neisser & Becklen, 1975, p. 491f]

Some of the hierarchical-prediction-error-minimization theory's strengths can be brought into view by applying the theory to the results of Neisser and Becklen's experiment. That experiment also enables us to see why it is necessary for the theory to introduce the notion of *precision optimization*; a notion which is crucial to its account of attention.

3. Attention and precision

About half of the visual information that was presented in Neisser and Becklen's experiment was consistent with the hypothesis that the filmed scene was one in which a hand game is being played. About half was consistent with the hypothesis that the scene was one in which a ball game is being played. Since these hypotheses are quite different, the information that is predicted by either one will tend not to be predicted by the other. The central claim of the prediction error theory is that the brain is only ever concerned with representing such top-down hypotheses, and such relations of unpredictedness: The content of experience is given by the top-down hypothesis that predicts the incoming signal most successfully (p. 48, 201); and the bottom-up signal tells us *only* about the extent to which the incoming information is unpredicted by our prior hypotheses: "only prediction error is propagated up through the system in a bottom-up fashion" (p. 47).

These features of the prediction error theory enable it to give a very natural account of why it should be "easy to attend to one sequential episode and ignore another, even when both are presented in haphazard overlap" (Neisser and Becklen, p. 491). The theory tells us that the brain does not need to first separate out the two overlapping streams of visual information,

[5] The fact that prediction error coding minimizes the coding of redundant information can also be formulated by saying that such codings maximize the amount of representational effort that is devoted to the encoding of information that is surprising (pp. 51–53). Redescribing the point from this opposite direction might seem to make the connection with attention more obvious, since we do indeed pay attention to surprising things. But the intuitive appeal of this recasting is somewhat dependent on equivocating between our normal first-person understanding of surprise, and a technical information-theoretic sense – 'surprisal' – in which the notion of surprise is here intended.

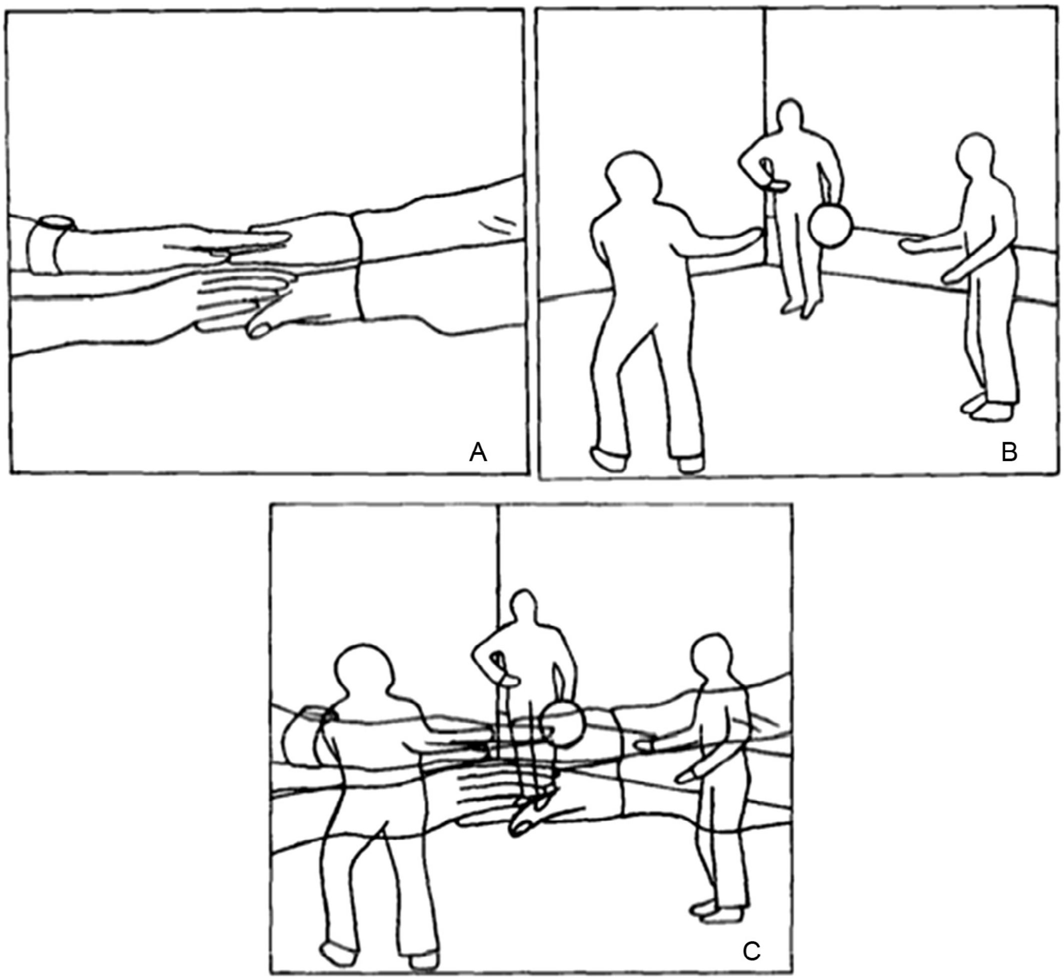

Fig. 1. Neisser and Becklen's (1975) 'Selective Looking' Experiment. Participants were presented with two overlapping streams of visual information (C), one from a hand clapping game (A) and one from a ball game (B). In some experimental conditions a different stream was presented to each eye, creating binocular rivalry. In other conditions both streams were presented together, to both eyes, by using a system of half silvered mirrors. Participants knew which of these conditions they were in.

in order to then build up a coherent hypothesis about the scene from which this information was derived. Instead the brain *starts out* with some coherent hypotheses, which it applies to the entire haphazard overlap of information. The different parts of that information can then be picked out on the basis of whether or not they are predicted by a given hypothesis. This provides a straightforward explanation for our ability to selectively attend to one or other of the scenes, despite the haphazardness of their overlap. The theory also gives an account of why it should be that simultaneous attention to both streams is 'demanding' or 'impossible' (Neisser and Becklen, p. 490). That difficulty can be explained as a consequence of the low prior probability that is assigned to the occurrence of overlapping scenes, and of the difficulty in deriving specific predictions about what should be expected on the hypothesis that two incompatible games are being played in the same place, and at the same time.[6]

The patterns of possible and impossible attentiveness that were observed by Neisser and Becklen can therefore be accommodated by the Bayesian prediction error minimization theory, and their accommodation is a relatively straightforward matter. It is nonetheless clear that a complete account of the attentiveness that is found in Neisser and Becklen's experiment requires that more complexity be added to the prediction-error theorists' account. The need for this complexity can be seen by noticing that, when the participants in Neisser and Becklen's experiment are attending to the hand game, they are expecting to see a hand game, and so will increase the probability that they assign to information that is consistent with the

[6] The proposed explanation draws on Hohwy, Roepstorff, and Friston (2008) analysis of binocular rivalry. Given that binocular rivalry and 'selective looking' are not strictly analogous phenomena, one might wonder whether the explanation is wholly adequate. Answering this question would take us beyond the main focus of this paper.

hypothesis that such a game is taking place. The information corresponding to the *other* game – the ball game – will therefore become the part of their sensory input that is not predicted by their currently favored hypothesis. That ball-game-related input will therefore generate the greatest prediction-error. And so the information pertaining to this *unattended* game will command the larger share of perception's representational resources. The prediction-error theory is therefore unable to say that the attended stream is to be identified with that stream to which the most representational resources are devoted, or that it is that stream which generates the greatest prediction error.[7] Nor could it be said that attention is paid to that stream of information which generates the *least* prediction error: if that were so then unexpected but salient stimuli would be attention-repelling, rather than attention-catching. The theory's account of attention therefore needs to be given in terms of something other than the magnitude of the prediction error signal.

Instead of accounting for attention by reference to the magnitude of the prediction error signal, the prediction error theorists' suggestion is that we should account for it by reference to the question of which prediction error signals are treated as being the most *precise*. Once this notion of precision has been introduced the explanation of attention is supposed to be straightforward: "attention is simply the process of optimizing precision during hierarchical inference" (Friston, 2009, p. 299).

Precision here refers to the inverse of a signal's *variance*: it is a measure of how much that signal fluctuates around its mean. It is therefore correlated with the extent to which the information given in that signal warrants the adoption of a fine-grained revision to one's prior hypotheses. To see this, suppose that instruments A and B both give a reading of y volts when applied to some particular cable. In order to comply with the present framework we should suppose that each instrument encodes this reading in the form of a prediction error signal. We should therefore suppose that our prior hypothesis is that the voltage on this cable is z volts ($\pm$ some n), and that the output given by each instrument is a representation saying, not that the actual voltage is y, but that our prior prediction of that voltage was x volts too high or low. We can also suppose that on this occasion the voltage being measured is indeed y volts exactly. Both instruments' readings would then be perfectly accurate. But if it is known that instrument A is more *precise* than instrument B – in the sense that the readings from it vary less when given a constant input – then the Bayesian reasoner who encounters the readings of these instruments should respond differently to them. After updating on readings that come from what is known to be a *precise* instrument, the probability density function of a rational thinker will be a relatively tight curve, with a relatively low dispersal around its mean. After updating on readings from what is known to be an *imprecise* instrument, the probability density function of a rational thinker should still be somewhat spread out. It is in this sense that the *precision* of a signal measures the extent to which the information given in that signal warrants the adoption of a fine-grained revision to one's prior hypotheses (it being these hypotheses that our probability density functions represent).

It is easy to believe that representations of precision might find application in the brain of a perceiving subject. When presented with an auditory signal indicating that a previously unexpected voice is coming from straight ahead of you, you should increase the probability that you assign to the hypothesis that there is someone speaking in that location, but if you know that such auditory signals give imprecise information about locations then conformity to Bayesian rationality requires that you commit to only an approximate specification of where that speaker is. If the same location is also indicated in vision – and if you know that vision is relatively precise about matters of location – then you may be warranted in committing to a location for the speaker that is more narrowly specified. If we suppose that the perceiving brain does indeed take account of such expected precisions then we can begin to account for our susceptibility to the illusion of ventriloquism, in which a visual cue overrides an auditory one to determine the perceived origin of a voice (Hohwy, 2013, chap. 7). This is certainly not the whole explanation for our susceptibility to the ventriloquism illusion (see Alais, Newell, & Mamassian, 2010; King & Walker, 2012), but if it is a part of that explanation then the brain must be representing the expected precision of the sensory signals that it is picking up, and it must be using information about that precision in building up its picture of the perceived scene. Similar things can be said about the interplay between vision and proprioception in the rubber hand illusion. It is plausibly because vision is expected to be especially precise that the visual signal dominates the proprioceptive signal (see Hohwy, chap. 5, Suzuki, Garfinkel, Critchley, & Seth, 2013). If some such explanation is correct then the brain must be using expectations of precision when building up its picture of the world.

In light of this need to take account of precision, the prediction error theorists claim that bottom up prediction errors will be weighted according to top down expectations of the precisions in these error signals, so that "effort is put in where the prediction error signal is expected to be precise" (p. 195, cf. p. 64). These expectations of precision should vary between different circumstances. On some occasions vision will be the most precise source of information. On other occasions – if the light is bad, or a fog descends – the more precise signal will be from audition. The weightings of these signals will therefore need to be adjusted if they are to be handled optimally:

> For a system that is able to optimize its precision expectations the pattern of effort will fluctuate in tune with the pattern of state-dependent noise.
>
> [p. 195]

[7] This is one illustration of why it is that Itti and Baldi's (2006) model of bottom-up attention fails to generalize to top-down attention.

In making such adjustments, we amplify those signals that are expected to best minimize overall prediction error, and dampen those that are expected to include a high level of noise. The prediction error theory therefore attempts to explain attention – and so "to provide alternative conceptualizations of [attention's] functional role in our overall mental economy" (p. 191) – by identifying it with the process by which these precision-expectation weightings are adjusted from one context to the next:

> The notion of precision was related directly to attention, such that attention is nothing but precision optimization in hierarchical inference.
>
> [Hohwy, p. 244, citing Feldman & Friston, 2010]

Unlike traditional theories of attention (such as that of Broadbent, 1958), and unlike some modern theories that inherit assumptions from these traditional sources (e.g., Lavie, Hirst, de Fockert, & Viding, 2004), the selectivity of attention is not taken by the prediction error minimization theorist to be the result of limitations in our capacity for perceptual processing, or for thought. Nor does that theory take this selectivity to be imposed by our need to act on one thing at a time (Allport, 1987; Wu, 2014). The theory 'provides an alternative conceptualization of attention's functional role' because attention does not introduce selectivity in order to cope with any sort of bottleneck in the amount of information that can be processed or acted upon. Attention's selectivity is instead understood to be a rational response to state-dependent fluctuations in the world's noise.

4. The variety of attention shifts

When, following Friston, Hohwy writes that "attention is nothing but precision optimization in hierarchical inference", his claim is explicitly a 'nothing but' claim. If true, it should apply with full generality. It should therefore apply to attention in all its forms. Psychologists have traditionally drawn distinctions between these forms. They have distinguished between feature-based and object-based attention; between endogenous and exogenous attention; and (what is not quite the same contrast) between voluntary and involuntary attention. Hohwy's 'nothing but' claim is supposed to apply to all these forms. His theory can most readily be applied to cases that involve the involuntary capture of our attention, especially when this is the result of strong and abrupt external stimuli, such as bangs or flashes.

Since precision is the inverse of variance, and since it is in the nature of noise to be a source of variance, any decrease in noise will lead to a corresponding increase in precision. Our environment is one in which sources of weak noise are relatively commonplace, and sources of strong noise relatively rare. It is therefore plausible to suppose that strong signals are relatively noise-free. If this assumption is granted then the capture of attention by strong signals can be taken to result from an automatic weighting of these, derived from a standing expectation to the effect that, in our environment, strong signals will tend to be more precise than weak ones (p. 197).

More substantive assumptions are needed in order for a similar account to be applied to the direction of our attention by arrows, by the line of another person's gaze, or by such things as pointing fingers and opening doors. These additional assumptions are not implausible. It is quite plausible that there has been a real correlation, in the world as we have experienced it, between the occurrence of such indicative stimuli and an elevated precision in the signals that they have indicated (p. 65). This correlation can be learned through a piece of second-order Bayesian reasoning, concerning the precision of signals in various precision-indicating contexts. A representation of this correlation can then be used as the source for expectations about the precision of subsequent signals (p. 195), and the process of prediction-error minimization can weigh those signals accordingly. Such empirically-derived expectations of precision can explain all of the cases in which attention is directed onto items that have been recognized as belonging to a category that has previously been perceived precisely. These will include cases in which attention is directed onto a thing by a precision-indicating stimulus that lies outside of the thing itself such as pointing arrows, fingers, or direction of gaze (Feldman & Friston, 2010; Hohwy, 2012).

Some psychologists refer to cases of this latter sort as cases of 'endogenous attention' (see, for example, McCormick, 1997). Others want to reserve the terms 'endogenous' and 'exogenous' to mark a distinction between those cases where the allocation of attention originates in some condition of the attending subject herself and those in which it originates in some condition of the world to which she attends. Whatever labels we apply to it, this latter distinction is certainly a real one. There are cases in which attention is not directed by any occurrence in the environment, but is instead directed by mental occurrences that are internal to the attentive perceiver. These might be occurrences of decision, emotion, strategy-adoption, or whimsy. Neisser and Becklen's experiment provides one example.

When presented with Neisser and Becklen's overlapping films, participants find that they can attend to either one, and that they can do so on the basis of arbitrary picking. Having started by attending to one film, it is an easy matter for them to switch attention to the other (Neisser and Becklen, 1975, p. 494). In allowing for such ad hoc switches of attention, the case of *overlapping* films (presented with half-silvered mirrors) is unlike the case of films that are presented one to each eye, under conditions of binocular rivalry. The latter case – in which the switching of attention is hard – is one that has been discussed frequently in the recent literature, and it is one that Hohwy himself discusses (pp. 19–23, 108f). Neisser and Becklen describe the contrast between these cases, referring to the conditions that create binocular rivalry as 'the dichoptic case', as follows:

> In binocular viewing [that is, in the half-silvered mirror scenario, with both eyes seeing *both* of the overlapping scenes] we find it easy to switch from one episode to the other. There is no feeling of effort or inertia as in the dichoptic case. The desired episode is available instantly when we look for it [...].
>
> [Neisser & Becklen, 1975, p. 494]

When attention gets switched from one of the overlapping films to the other, in the way that Neisser and Becklen find to be easy, this switch can take place without the occurrence of any external cue and so, a fortiori, it can take place without any external cue that has a known correlation with enhanced precision: it is simply that "The desired episode is available instantly when we look for it". When we decide to switch our attention from one scene to the other, there must be some sort of top-down signal that brings this switch about. Hohwy claims that this is a signal that changes the weightings of the various bottom-up prediction error signals, by adjusting the relative gain that is applied to them. He also makes a claim about the content of this gain-adjusting signal, and this latter claim is one that he takes to be crucial. "The crucial question", he says, "is on what basis the mechanism sets the gain on prediction error" (p. 194). Hohwy's reconceptualization of attention is given in his answer to that 'crucial question', which is that the gain-adjusting signal should be characterized as the representation of an expectation concerning the *precision* of the prediction-error signals. It is against this part of Hohwy's theory that we wish to raise an objection. Our objection is based on the fact that we can see no grounds for thinking that the proposed characterization of the top-down gain-adjusting signal is correct.

To see the root of this objection, remember that the precision of an incoming signal is a statistical feature of that signal, over some interval. A signal's precision will vary depending on the context in which that signal is being received, but once a context has been fixed that signal's precision is an objective feature of it. It is these *objective* precisions that attention attempts to track, on Hohwy's theory of its function: "Precision expectations must be based in an internal model of precisions *in the world*" (p. 65, emphasis added).[8] These precisions 'in the world' are not influenced by the goal, task, or mood of the perceiving subject. A visual signal's precision will be objectively lower at night than it is during the day, and objectively lower in conditions of fog than conditions of clarity. The visual system will very probably have formed an expectation of this, having learned about the contexts in which such signals tend to be more and less precise. That expectation might issue in an attention-switching signal of the sort that Hohwy describes. But in Neisser and Becklen's experiment the precisions *in the world* are equal. The signal coming from the attended film is no more precise than is the signal coming from the film that one chooses to ignore. As Neisser and Becklen remark: "What defines one episode and distinguishes it from the other is not [...] its clarity" (Neisser & Becklen, 1975, p. 491). The prior experiences of the subject give her no reason to expect otherwise. And so the signal that shifts this subject's attention from the hand-game to the ball-game cannot be derived from any process that is tracking the expectations of precision in the world. This signal cannot be accounted for by a theory that "reduces attention to a simple matter of learning regularities of precision in nature" (p. 205). Even if one were to insist that the content of the gain-adjusting signal *is* an expectation of precision, the process by which such an expectation was formed must be a process that proceeds by non-Bayesian means. There must therefore be some place at which Hohwy's claim to completeness breaks down, and where the brain of the voluntarily attending subject is doing something in addition to the Bayesian minimization of prediction error.

5. Active inference

The argument above does not yet take account of an important complication, coming from Hohwy's theory of action. That theory increases the complexity of the problem that we have just identified, without – we contend – doing anything to address the ultimate source of it.

In order for actions to be subsumed within the prediction-error theory they must be explained as attempts to minimize the prediction errors that are generated by our top-down hypotheses. The minimization of such errors is, Hohwy tells us, all the brain ever does (p. 7). But his theory allows there to be two different ways in which the brain does it. Our discussion so far has been concerned with inferences in which the brain minimizes prediction error by updating its top-down hypotheses in response to its incoming signal. The theory also allows for cases in which the brain minimizes prediction error by changing that incoming signal to match its top-down hypotheses. The most straightforward way in which this change can be made is by intervening in the world: if there are any prediction errors generated by the carpenter's expectation that his struck nails will be flush with the board, those errors can be minimized by increasing the force with which he wields his hammer. When a carpenter minimizes his prediction errors in this way, the prediction error theorist says that he engages in 'active inference'. Action is thereby understood to involve the same principles of Bayesian prediction-error minimization as are involved in perception.

By itself the introduction of active inference does not give an explanation for purely mental acts, such as the act of directing one's attention. This is just because the agent who acts mentally does not intervene in the world. The subject in Neisser and Becklen's experiment remains where she is, and the world goes on, just as predictably or unpredictably as before. The only thing that changes is the direction of her attention. Again it is considerations of *precision* that are supposed to address this. The theory claims that – although the attentive subject does not change the world – she does sample that world differently, on the basis of changing her expectations of precision. The mental act of attending is therefore:

[8] We can also look to Feldman and Friston (2010): "the precision of sensory signals depend on states of the world. This means that optimizing precision entails optimizing inferred states of the world that affect the precision or uncertainty about our sensations" (2).

> a slightly unusual instance of action because the way we change our relation to the world is to increase the sensory gain in one region of space. The difference is then that the active inference is driven not by *which* prediction error we expect, but by the *precision* of the expected prediction error.
>
> [p. 198]

To see how this 'slightly unusual' case is supposed to work, recall, from Section 3, that our brains seem to make use of some hypotheses that are not simply about the world: they seem also to make use of hypotheses about the precision of the signals that they are receiving from that world. The prediction errors generated by hypotheses of either sort can be minimized through inferences that are active, as well as through those that are not: we can change our beliefs about the precision of the prediction error signals that we are receiving, or can intervene in various ways that will *make* those signals more or less precise. To increase the precision of the visual signal, for example, we might turn on a flashlight, or move ourselves nearer to the window. Since *attention* is to be understood as the optimization of the precision-weighting of our prediction errors, and since *action* is to be understood as active inference, the action of voluntarily attending needs to be understood as some such precision-changing movement.

The subject in Neisser and Becklen's experiment does not intervene on the precision of her incoming signals by the use of a flashlight, any more than she minimizes prediction error by wielding a hammer. Hohwy's suggestion is that she nonetheless does perform an internal act that is somewhat analogous to a flashlight-based manipulation of precision. Her act of attention-switching is to be construed as actively bringing her expectations of precision into line with the actual precisions of the prediction-error signals that she is receiving, via an intervention on those signals. The interventions in question are taken to be manipulations of sensory gain, so that "the way we change our relation to the world is to increase the sensory gain in one region of space" (p. 198).

If these manipulations of sensory gain are to be understood in the way that Hohwy suggests then the top-down signals that produce them must be interpreted as encoding expectations of precision. Those expectations have an important status in Hohwy's theory since they are able to be self-fulfilling (p. 90). Their mechanism of self-fulfillment is a simple one, depending on a point that we have already seen: weak noise will be more commonplace than strong noise (in the brain as it is in the world), and so those signals on which the gain has been increased will be less likely to get mixed up in background noise. Those signals will therefore show less variance. They will, as an immediate consequence, be more precise. If expectations of high precision can lead to an increase in gain, those expectations can make themselves true.

Because the expectation of precision in one stream of information can cause that stream to become more precise, the *truth* of such expectations is unproblematic. But we contend that our original problem remains. The theorist of prediction error minimization has said that any act of voluntary attention involves an adjustment being made to the gain on an incoming signal. More specifically (and more controversially), they have claimed that such an act involves an increase of the gain on a prediction-error signal. Most controversially (and most 'crucially') they have said that the cause of this increase in gain is a signal having as its content the hypothesis that such an increase of gain would lead to an increase in the precision of that signal. It is this last claim that is the focus of our objection, which continues to be that there is nothing in virtue of which the gain-adjusting signal can claim to have *this* expectation as its content. If attention-based manipulations of sensory gain are to be understood in the way that Hohwy is suggesting then the expectations that produce them must be interpreted as expectations of precision. Even once the notion of active inference has been introduced, Hohwy is unable to give an account of why this should be the correct interpretation of those signals' content. The signals cannot get that content by being tuned to evidence of its truth: none of the exogenous information that is available to the subject in Neisser and Becklen's experiment indicates that signals coming from the hand game will be any more precise than signals coming from the ball game, and yet there *is* a top-down signal that shifts the subject's attention from one to the other.

5.1. Hypotheses with non-indicative content

The basic claims of the precision error minimization framework imply that the content of every top-down signal must be an hypothesis, and the content of every bottom-up signal must be that some such hypothesis is somehow in error. To be clear about this framework's account of voluntary attention (and about our objection to it), we must distinguish between different sorts of hypotheses, all of which are said to be generating predictions that are being tested in the brain of a voluntarily attentive subject. Of these there are at least three. The first are first-order categorical hypotheses, saying that a ball game is taking place (in some way), or that a hand game is taking place (in some way). The second – introduced by the considerations that we reviewed in Section 3 – are hypotheses concerning the precisions of the various signals that carry information about the prediction errors made by those first hypotheses. The third – introduced by the theory of active inference, as applied to the case of attention – are *conditional* hypotheses, concerning the effects on those precisions of manipulations of the gain on these prediction-error signals.

In this theory of active inference, actions are understood to be encoded as the antecedents in hypotheses that have the form of a counterfactual conditional saying that, if the action in question were to be performed, then variable V would be expected to have value y ($\pm n$). In acting, we gather data about the errors in these counterfactual hypotheses. We thereby minimize the prediction errors in our overall picture of the world (pp. 81–89, Friston, Daunizeau, Kilner, & Kiebel, 2010). When we ask about the top-down signal that implements an act of voluntary attention, it is an hypothesis of this third sort that is said to give that signal's content.

It has long been known that the introduction of counterfactual conditionals to a Bayesian calculation leads to a paradox if we fail to distinguish between the probability of a conditional and a conditional probability: We must distinguish between the probability that some whole conditional sentence is true, and the probability of that sentence's consequent, conditional on its antecedent (Hájek & Hall, 1994; Lewis, 1991). It must therefore be one thing to update the probability that we assign to the claim that a signal's precision would be improved by a manipulation of gain, but it must be a different thing to update the probability that we would assign to an increase in that precision, if gain were being manipulated. When voluntary attention is understood as an instance of active inference, the top-down signal by which sensory gain is modulated purports to have the first of these as its content. That signal's content is supposed to be that sensory precision would be improved by a manipulation of gain.

It is this that makes the case of voluntary attention quite unlike the cases of endogenous and exogenous attention that were previously discussed, in which the gain on prediction error was set on the basis of precision expectations that had been derived from the observation of regularities in the world. In those cases attention was selective because some signals were expected to be more precise than others. Because those expected precisions vary from context to context, there was a need to select whichever signal is expected to be precise in the present context, in order that the gain on that signal can be boosted accordingly. But the probability of the conditional saying that an incoming signal would become more precise if its gain were increased is uniformly high. It is high whichever signal it pertains to. No top-down signal that is produced at the time when an act of attention occurs has any more claim to this conditional content than do the signals existing before or after that act. None of these signals can claim to sustain a personal-level thought with the relevant subjunctive content (since the voluntarily attending subject may not have such thoughts). And all of these signals can claim to have that subjunctive content as a part of the information that they subpersonally carry, since the content in question is sufficiently close to being a truism: one would expect increased precision *whichever* object one's attention were directed towards, regardless of the actual precision in the incoming signal.[9] Such counterfactual precision expectations cannot be what set the gain on prediction error because there is no way of adjudicating between them; acting on any of them will lead to enhanced precision. At the moment when one attends to *x* rather than *y* one expects that the precision of signals pertaining to *y* would be enhanced by an increase in their gain, just as one expects that the precision of signals pertaining to *x* would be. These counterfactual expectations are on a par.

The preceding considerations leave the theorist of prediction error minimization without any account of voluntary attention's selectivity. And without giving an account of this, the prediction error minimization theory of voluntary attention becomes little more than a truism. The theory tells us that shifts in attention occur because we have standing expectations of the form 'precision *y* would be expected if an action of enhancing the sensory gain were to be performed', where the precision that would counterfactually be expected is higher than the precision that is actually expected. But since the theory takes this action of enhancing the sensory gain to *be* the act of attending, this precision expectation amounts to no more than the claim that we would enhance the precision of the signal coming from whatever we attended to. On the assumption that enhanced precision results in perceptual enhancement, this amounts to no more than the truism that whatever we decide to attend to will, to some degree, be enhanced perceptually. That conclusion is one that many theorists of attention could agree to, whether or not they are operating with a predictive coding framework.[10] They would not take it to be an explanatory account of top down attention, because it does not provide any analysis of attention's voluntary selectivity.

5.2. Shifting to active inference

The issue that we are raising can be seen as one part of a more general challenge for the predictive coding framework, arising from its failure to account for the shift from perceptual to active inference. Both modes of inference can reduce prediction error. If the minimization of such errors is the only goal at which the brain aims then we should prefer whichever mode of inference reduces them most effectively. And since the world in which we act is more recalcitrant than our internal model of it, perceptual inference would seem to have a standing advantage. The shift back, from active inference to perceptual inference might plausibly happen by default, but shifts from perceptual to active inference require an explanation. Some psychological event must bring them about. When a shift of this sort gets brought about by a top-down signal, the Hohwyian theory must say that that signal encodes an hypothesis the content of which has been arrived at by Bayesian means. The challenge is to specify such an hypothesis.

This challenge might not seem to have anything in particular to do with attention. It does because Hohwy's analysis of voluntary attention builds active inference into its account: any voluntary decision to attend must take the form of an active inference.

Hohwy attempts to take a step towards a solution to the problem of how we switch from perceptual to active inference. Combining the idea about attention with a suggestion about proprioception, he writes that:

[9] There will, unavoidably, be a ceiling on the quantity by which the precision can be increased, but this, again, is not due to the states of the world. It is due to limits on the amplification of the signals in the cortex.

[10] Those who think that perceptual enhancement is not essential to attention include Ganson and Bronner (2012) and Watzl (2011). Hohwy (2012) writes that in the case of pure volitional attentional switching to low contrast gratings, "the prediction is then that there will be less enhancing effect of such pure endogenous [volitional] attention since the high expected precision expectation (increased baseline) in this case is not induced via a learned causal regularity linking strong signal cues to strong signal targets" (p. 11).

> One intriguing proposal is that this mechanism is attentional, focused on the precisions of proprioceptive input (Brown, Adams, Parees, Edwards, & Friston, 2013). Briefly put, action ensues if the counterfactual proprioceptive input is expected to be more precise than actual proprioceptive input, that is, if the precision weighted gain is turned down on the actual input. This attenuates the current state and throws the system into active inference.
>
> [p. 83]

This proposal will not help in the present case, in part for reasons that we have already discussed. The counterfactual precision expectation for the ball-game-related signal will be higher than the actual precision expectation for that signal, but this much will be true for any counterfactual precision expectations that count as 'self-fulfilling prophecies'. The counterfactual hypothesis pertaining to the ball-game-related-signal has nothing to specially qualify it as the content of the signal which drives the shift in attention.

In addition, it is not clear what, if anything, is supposed to generate the relevant proprioceptive prediction errors in the case of voluntary attentional shifts. Recall that the action in question is to adjust the sensory gain on prediction error for a given region of space. This is not a matter of physical movement, and so cannot involve proprioceptive feedback. Yet if there is no source of prediction error feedback signal, then there cannot be any precision expectations pertaining to the 'action' either, and so action cannot be initiated on this proposal. We therefore remain skeptical that the problem of accommodating voluntary attention – and the related problem of switching from perceptual to active inference – can be surmounted using only resources that are internal to the theory of prediction error minimization. That theory's central claims are (1) that the representations coming from the top down only ever carry content of the form 'Variable v is expected to have value x ($\pm n$)', or else of the form 'Value y ($\pm n$) would be expected if action a were to be performed'; (2) that the representations coming from the bottom up only ever carry content with the form 'Variable v has an actual value that is z units higher (or lower) than was predicted'; and (3) that the representations that regulate the interactions between these top-down and bottom-up information streams are either representations of the probability of the bottom-up signals, conditional on various other top-down hypotheses, or else are representations that carry content of the form 'These estimates of the error in your prediction as to the value of variable v have a precision of $\pm n$'. Each of these claims places a restriction on permissible content. Those restrictions are not optional to the theory. They are crucial to the parsimony that is supposed to be one of that theory's main advantages but – since the content of signals is restricted in every case to *indicative* content – it is doubtful that all mental content can be shoehorned into one or other of these categories. The theory has a built-in resistance to non-indicative content: Content which makes no claim about the way the world is cannot be assessed for truth; and cannot have a probability assigned to it. Nor can a probability assignment be conditional on it. Our objections stem from this difficulty.

We have argued that the decision to attend to the hand-game over the ball-game cannot be analyzed by reference to probabilities that are assigned to values for the actual precision of the signal, nor by reference to the probability that is assigned to the counterfactual claiming that that precision would be enhanced if gain were to be increased. The shift of attention that results from the decision to attend cannot be brought about by a signal that has an hypothesis concerning either of these probabilities as its content.[11] Other non-indicative contents are similarly resistant to Bayesian handling: neither imperative content (Hall, 2008; Klein, 2007; Martínez, 2010), nor interrogative content (Eilan, 1998; Koralus, 2014), can be straightforwardly accommodated by the prediction error minimization framework.[12] That framework is fundamentally committed to the idea that the sources of top-down signals are hypotheses that get updated via a process of Bayesian inference, but Bayesian reasoning can make no use of non-indicative content.

It may be suggested that recent advances in statistical causal modeling provide a solution to this objection. The "do" operator, suggested by Pearl (2000; 2010), for instance, can be used to set the value for a set of variables. The variables of interest here are the gain on the prediction error units corresponding to features of the scene to which we intend to direct our attention (features of, say, the ball game). We could thus arrive at the relevant counterfactual by conditioning the expected precision of the signals arriving from the ball game on the information that the gain on these signals is set to a higher value. And this counterfactual will typically be true: given that the gain on the relevant units has been set to a higher value, it will typically be true that the precision of the signals arriving to these units will increase.

In order to see that this suggestion fails to address our objection, recall the earlier point that we should distinguish between the probability that the whole counterfactual conditional is true, and the probability that its consequent is true, given its antecedent. Both scenes in Neisser and Becklen's experiment are on a par with respect to the latter: in both cases, given an increase on the relevant set of gains, it will typically be the case that the precision of the corresponding signals

[11] It is also doubtful that desires more generally can be so reduced. On the predictive coding account, desires are analyzed in terms of active inference: to have a desire just is to be testing the errors generated by a counterfactual, and to act on a desire just is to minimize the error generated by that counterfactual, by changing the world in some respect (p. 89). But this reduces desire to belief; a reduction that leads, as Lewis showed, to inconsistency (Lewis, 1988, 1996). While others have argued that there are fixes available to avoid this problem, we take it that such fixes are not available to the evidential Bayesian, and so that predictive coding cannot incorporate them.

[12] Shea (2013) suggests that active inferences employ imperative content by virtue of having a different direction of fit from perceptual inferences, but Colombo (2016) gives an argument to the effect that, on the predictive coding theory, cognitive and conative states must have the same direction of fit. We thank an anonymous reviewer for drawing our attention to this literature.

would also increase. What the prediction-error theorist owes us is an account of *why the relevant counterfactual*, *as a whole*, *is selected*: They need to provide an account of something that makes it the case that we increase the gain on the ball game-related set of features, rather than increasing the gain on the hand clapping game-related set of features. Absent such an account, the selective shift of attention goes unexplained.[13]

A final objection is that Hohwy does have the resources to accommodate non-indicative content because, following Friston and colleagues, he 'reduces' utility to Bayesian inference. (We would like to thank an anonymous reviewer for pressing us to clarify this issue.) Such a reduction would be based on the formal result that predictions of optimal behaviour using decision theory can be achieved without using reward or cost functions. Instead, optimal behaviour can be reconstrued in terms of active inference, where the explicit goal is not to minimize cost or maximize utility but instead to minimize prediction error.[14] On this picture, reward or cost functions are replaced by 'prior beliefs about state transitions'. State transitions should be understood as action policies to be undertaken by the agent in specific contexts, such as 'seek shelter when it rains'. Minimizing prediction error with respect to these beliefs will lead to optimal behaviour (maximizing utility): "if I believe that I will seek shelter when it rains, then I will behave optimally, provided I act to fulfill these beliefs" (Friston, Samothrakis, et al., 2012, p. 524). While we are dubious of the prospects of such an elimination for much the same reasons put forth by Gershman and Daw (2012, section 5.1), supposing that this commitment is taken on board leaves our criticisms unchanged. It does not solve the problem of how active inference is initiated over perceptual inference. It is not clear what sort of action policy, or hypothesis about state transitions, could account for the timing and manner in which shifts of voluntary attention occur. It cannot be an action policy of the form 'attend to what you want', or 'attend to what you decide to attend to' because this would be circular. The action policy cannot rely on the very notions it seeks to account for. If it does not rely on these notions however, it is unclear how is it initiated.

Indeed, such a reduction is ostensibly the source of the problem: if all mental states are a species of hypothesis about the world, and the objective is to minimize overall prediction error, then the surest way to achieve this objective is to find a dark room and stay there instead of engaging in active inference. This is known as the 'dark room' problem. While theorists have proposed solutions to the 'dark room' problem based on general evolutionary considerations (Friston, Thornton, & Clark, 2012; Hohwy, 2013), these general considerations fall short of explaining how *in particular instances* the voluntary switch from perceptual to active inference is accomplished. We have already seen that the solution Hohwy considers, in terms of precision optimization, fails at least in the case of voluntary attention we have been discussing. If attention is still to be identified with the optimization of expected precisions then replacing utilities with first-order hypotheses about state transitions does not do anything to alter our second-order hypotheses concerning expected precisions. Signals will continue to be equally precise or imprecise.[15]

6. Conclusions

Hohwy tell us that his "account of attention is interesting because it reduces attention to a simple matter of learning regularities of precision in nature" (p. 205). That account fails to apply to cases in which it is not regularities in nature that determine the way in which attention is allocated. There are instances of voluntary attention where the gain cannot be set by expectations of precision (whether or not these expectations are 'learnt from nature'). When the relevant expectation is a self-fulfilling prophecy, then the gain must be set by something other than precision, because all such expectations will be on a par.

The objections that we have raised above can be thought to arise from the fact that there are some contents – in particular, some non-indicative contents – that cannot be expressed by representations having any of the forms available to the predictive coding framework. It seems clear that such non-indicative content plays a role in the allocation of attention. Since it lacks the representational resources to encode these contents, the prediction-error minimization theory is unable to account for those occasions on which attention is allocated on the basis of states that essentially involves content of this sort. This undermines that theory's claim to completeness. It does not deny that the theory casts valuable light on a considerable number of attention-related phenomena.

[13] What is more, it would not do to force an analogy between the present case of voluntary attention, and a case of causal inference by (semi) self-supervised systems. For the analogy fails. In the case of causal inference, the system has more evidential support for a particular causal model based on its previous interactions with the environmental regularities of interest, whether through "observations" or "interventions". And to test *that* model, it activates a counterfactual that follows from it such as to further its evidential support. In the case of voluntary shifts of attention, however, there is no environmental regularity that would "favor" attending to the unattended scene, as opposed to the currently attended one. In other words, there is no model of the environment that would increase the probability of increasing the gain on the unattended game (or activate the relevant "do" operator). If anything, the system is more justified in sticking with the current situation.

[14] More exactly, "in active inference, actions and beliefs about hidden states minimize a variational free energy bound on the (negative log) marginal likelihood of observed states, that is, they maximize the marginal likelihood" (Friston, Samothrakis, & Montague, 2012, p. 523).

[15] Note that while Clark (2013) does not endorse the eliminativist position of Hohwy and Friston, where goals and reward signals are converted into expectations, he also does not endorse a position where utility and probability are represented separately. He speculates that "instead, it seems likely that we represent the very events over which probabilities become defined in ways that ultimately fold in their personal, affective, and hedonic significance" (p. 200). While it is beyond the scope of this paper to evaluate this claim, note that insofar as Clark endorses an explanation of attention solely in terms of precision optimization, our criticisms continue to apply to his view.

Acknowledgments

Madeleine Ransom and Sina Fazelpour are both supported by the Joseph Armand Bombardier Canada Graduate Scholarship Program. Christopher Mole's research is funded by the Social Sciences and Humanities Research Council of Canada, Grant No.: 435-2015-0321.

The authors would like to thank Andy Clark, Adrian Downey, Dustin Stokes, Jona Vance, and the referees for this journal.

References

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving, 23*(1), 3–38. http://dx.doi.org/10.1163/187847510x488603.

Allport, D. A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer & H. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 395–419). Lawrence Erlbaum Associates Inc.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception, 1*(4), 371–394. http://dx.doi.org/10.1068/p010371.

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise? *Frontiers in Psychology, 4*. http://dx.doi.org/10.3389/fpsyg.2013.00907.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Bower, T. G. R. (1971). The object in the world of the infant. *Scientific American, 225*(4), 30–38.

Broadbent, D. (1958). *Perception and communication*. Oxford: Pergamon Press.

Brown, H., Adams, R., Parees, I., Edwards, M., & Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing, 14*(4), 411–427. http://dx.doi.org/10.1007/s10339-013-0571-3.

Carlin, B. P., & Louis, T. A. (2000). Empirical Bayes: Past, present and future. *Journal of the American Statistical Association, 95*(452), 1286. http://dx.doi.org/10.2307/2669771.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*(13), 1484–1525.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181–204. http://dx.doi.org/10.1017/s0140525x12000477.

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

Colombo, M. (2016). Social motivation in computational neuroscience. Or if brains are prediction machines, then the Humean theory of motivation is false. In J. Kiverstein (Ed.), *Routledge handbook of philosophy of the social mind*. Routledge (in press).

Colombo, M., & Wright, C. (2015). Explanatory pluralism: An unrewarding prediction error for free energy theorists [Preprint] Available at<http://philsci-archive.pitt.edu/11783/>.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation, 7*(5), 889–904. http://dx.doi.org/10.1162/neco.1995.7.5.889.

Doya, K. (Ed.). (2007). *Bayesian brain: Probabilistic approaches to neural coding*. Cambridge, Massachusetts: MIT Press.

Eilan, N. (1998). Perceptual intentionality, attention and consciousness. In A. O'Hear (Ed.), *Royal Institute of Philosophy Supplement 43: Current issues in philosophy of mind*. Cambridge, UK: Cambridge University Press.

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

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815–836. http://dx.doi.org/10.1098/rstb.2005.1622.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology, 4*, e1000211. http://dx.doi.org/10.1371/journal.pcbi.1000211.

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences, 13*(7), 293–301. http://dx.doi.org/10.1016/j.tics.2009.04.005.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics, 102*(3), 227–260. http://dx.doi.org/10.1007/s00422-010-0364-z.

Friston, K., Samothrakis, S., & Montague, R. (2012). Active inference and agency: Optimal control without cost functions. *Biological Cybernetics, 106*(8–9), 523–541. http://dx.doi.org/10.1007/s00422-012-0512-8.

Friston, K., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Frontiers in Psychology, 3*. http://dx.doi.org/10.3389/fpsyg.2012.00130.

Ganson, T., & Bronner, B. (2012). Visual prominence and representationalism. *Philosophical Studies, 164*(2), 405–418. http://dx.doi.org/10.1007/s11098-012-9853-3.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall.

Gershman, S. J., & Daw, N. D. (2012). Perception, action and utility: The tangled skein. In M. I. Rabinovich, K. J. Friston, & P. Varona (Eds.), *Principles of brain dynamics: Global state interactions* (pp. 293–312). Cambridge, MA: MIT Press.

Hájek, A., & Hall, N. (1994). The hypothesis of conditional construal of conditional probability. In E. Eells & B. Skyrms (Eds.), *Probability and conditionals: Belief revision and rational decision* (pp. 75–113). Cambridge: Cambridge University Press.

Hall, R. (2008). If it itches, scratch! *Australasian Journal of Philosophy, 86*(4), 525–535. http://dx.doi.org/10.1080/00048400802346813.

Harrison, L., Bestmann, S., Rosa, M. J., Penny, W., & Green, G. G. R. (2011). Time scales of representation in the human brain: Weighing past information to predict future events. *Frontiers in Human Neuroscience, 5*, 37. http://dx.doi.org/10.3389/fnhum.2011.00037.

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

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

Hohwy, J. (2016). The Self-Evidencing Brain. *Noûs, 50*, 259–285. http://dx.doi.org/10.1111/nous.12062.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition, 108*(3), 687–701. http://dx.doi.org/10.1016/j.cognition.2008.05.010.

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *WIREs: Cognitive Science, 2*(5), 580–593. http://dx.doi.org/10.1002/wcs.142.

Itti, L., & Baldi, P. (2006). Bayesian surprise attracts human attention. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.). *Advances in neural information processing systems* (Vol. 18, pp. 1–8). Cambridge, MA: MIT Press.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology, 4*, e1000209. http://dx.doi.org/10.1371/journal.pcbi.1000209.

King, A. J., & Walker, K. M. (2012). Integrating information from different senses in the auditory cortex. *Biological Cybernetics, 106*(11–12), 617–625.

Klein, C. (2007). An imperative theory of pain. *Journal of Philosophy, 104*(10), 517–532. http://dx.doi.org/10.5840/jphil2007104104.

Koralus, P. (2014). The erotetic theory of attention: questions, focus and distraction. *Mind and Language, 29*(1), 26–50. http://dx.doi.org/10.1111/mila.12040.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133*(3), 339–354. http://dx.doi.org/10.1037/0096-3445.133.3.339.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 20*(7), 1434. http://dx.doi.org/10.1364/josaa.20.001434.

Lewis, D. (1988). Desire as belief. *Mind, XCVII*(387), 323–332. http://dx.doi.org/10.1093/mind/xcvii.387.323.
Lewis, D. (1991). Probability of conditionals and conditional probabilities II. In F. Jackson (Ed.), *Conditionals*. Oxford: Oxford University Press.
Lewis, D. (1996). Desire as belief II. *Mind, 105*(418), 303–313. http://dx.doi.org/10.1093/mind/105.418.303.
Martínez, M. (2010). Imperative content and the painfulness of pain. *Phenomenology and the Cognitive Sciences, 10*(1), 67–90. http://dx.doi.org/10.1007/s11097-010-9172-0.
McCormick, P. A. (1997). Orienting attention without awareness. *Journal of Experimental Psychology: Human Perception and Performance, 23*(1), 168–180. http://dx.doi.org/10.1037/0096-1523.23.1.168.
Moore, J. B., & Weiss, H. (1979). Recursive prediction error methods for adaptive estimation. *IEEE Transactions on Systems, Man and Cybernetics, 9*(4), 197–205. http://dx.doi.org/10.1109/tsmc.1979.4310182.
Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica, 27*, 84–92.
Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology, 7*(4), 480–494. http://dx.doi.org/10.1016/0010-0285(75)90019-5.
Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.
Pearl, J. (2010). Causal inference. In *NIPS Causality: Objectives and Assessment* (pp. 39–58).
Rao, R. P. N., & Ballard, D. H. (1999). *Nature Neuroscience, 2*(1), 79–87. http://dx.doi.org/10.1038/4580.
Rao, R. P., Olshausen, B. A., & Lewicki, M. S. (2002). *Probabilistic models of the brain: Perception and neural function*. Cambridge, Massachusetts: MIT Press.
Richards, J. (2010). *The art and science of intelligence analysis*. Oxford: Oxford University Press.
Shea, N. (2013). Perception vs. action: The computations may be the same but the direction of fit differs: Commentary on Clark. *Behavioral and Brain Sciences, 36*(3), 228–229. http://dx.doi.org/10.1017/s0140525x12000477.
Stefanics, G., Kremlácek, J., & Czigler, I. (2015). Visual mismatch negativity: A predictive coding view. *Frontiers in Human Neuroscience, 8*, 1–19.
Suzuki, K., Garfinkel, S. N., Critchley, H. D., & Seth, A. K. (2013). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia, 51*(13), 2909–2917. http://dx.doi.org/10.1016/j.neuropsychologia.2013.08.014.
Vance, J. (2015). Review of Jakob Hohwy's. The predictive mind. Notre Dame philosophical reviews Published online Feb 27th, 2015. Retrieved February 16th 2016<https://ndpr.nd.edu/news/55878-the-predictive-mind/>.
Watzl, S. (2011). Attention as structuring of the stream of consciousness. In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: Philosophical and psychological essays*. Oxford: Oxford University Press.
Wu, W. (2014). *Attention*. Routledge.