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.