Brain and Cognition 112 (2017) 58–63

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Prediction error minimization: Implications for Embodied Cognition and the Extended Mind Hypothesis

Leon de Bruin [a,*], John Michael [b]

[a] Department of Philosophy, Radboud University Nijmegen, Erasmusplein 1, 6500 HD Nijmegen, The Netherlands
[b] Department of Cognitive Science, Central European University Budapest, Frankel Leó út 30-34, Budapest 1023, Hungary

ARTICLE INFO

Article history:
Received 1 August 2015
Revised 18 December 2015
Accepted 28 January 2016
Available online 5 February 2016

Keywords:
Prediction error minimization
Embodied Cognition
Extended Mind Hypothesis

ABSTRACT

Over the past few years, the prediction error minimization (PEM) framework has increasingly been gaining ground throughout the cognitive sciences. A key issue dividing proponents of PEM is how we should conceptualize the relation between brain, body and environment. Clark advocates a version of PEM which retains, at least to a certain extent, his prior commitments to Embodied Cognition and to the Extended Mind Hypothesis. Hohwy, by contrast, presents a sustained argument that PEM actually rules out at least some versions of Embodied and Extended cognition. The aim of this paper is to facilitate a constructive debate between these two competing alternatives by explicating the different theoretical motivations underlying them, and by homing in on the relevant issues that may help to adjudicate between them.

1. Introduction

Over the past few years, the prediction error minimization (PEM) framework has increasingly been gaining ground throughout the cognitive sciences. PEM essentially treats the brain as a probabilistic inference system, which is hierarchically organized in levels, and attempts to predict the input it receives by constructing models of the possible causes of this input (Clark, 2013; Friston, 2010; Hohwy, 2014). The main aim of the system is to minimize the 'prediction error', i.e. the discrepancy between the predicted and the actual input.

A key issue dividing proponents of PEM is how we should conceptualize the relation between brain, body and environment. Clark (2013) advocates a version of PEM which is committed to Embodied Cognition and the Extended Mind Hypothesis. He argues that some bodily and extended processes may qualify as constituting cognition and thereby reduce complexity for the brain, making it possible to interact with and exploit some features of the environment without representing them. Hohwy (2014), by contrast, presents a sustained argument that PEM actually rules out at least some versions of Embodied Cognition and the Extended Mind Hypothesis. Specifically, he argues that PEM in fact entails a boundary between cognitive systems and their bodies/environments, and that the concept of a 'Markov blanket' provides a principled basis for specifying that boundary.

The aim of this paper is to investigate how PEM constrains the relation between brain, body and environment, and what it implies for Embodied Cognition and the Extended Mind Hypothesis.

The paper has the following structure. In the next section (Section 2) we discuss the basic concepts and claims of PEM. In Section 3 we spell out the differences between Clark (2013) and Hohwy (2014) with respect to what PEM implies for Embodied Cognition and the Extended Mind Hypothesis. In Section 4 we trace these differences back to five fundamental issues, and use this as a basis for identifying means of adjudicating between the two approaches. In Section 5, we conclude by pointing out some directions for future research.

2. Prediction error minimization: a Primer

The basic idea behind PEM is that the brain is a kind of prediction machine: its goal is to anticipate incoming sensory, proprioceptive and interoceptive input as well as it can.[1] In order to achieve this, it constructs models of the possible causes of those inputs. These models generate predictions about likely inputs at any given time, which can then be compared to actual inputs. If the discrepancy between predicted and actual inputs – i.e. the

* Corresponding author.
E-mail address: lcdebruin@gmail.com (L. de Bruin).

[1] For the sake of simplicity, we will focus on *sensory* input unless otherwise noted.

http://dx.doi.org/10.1016/j.bandc.2016.01.009
0278-2626/

prediction error – is small, then there may be no need to revise the model that gives rise to the prediction. If, on the other hand, the prediction error is large, then it is likely that the model fails to capture the causes of the inputs, and therefore must be revised. In this sense, the brain is not concerned with coding input per se but only *unexpected* input.[2]

The models of the world that enable the brain to predict inputs are organized in a *hierarchy*. At the lowest level of the hierarchy, neural populations encode such features as surfaces, edges and colors. At a hierarchically superordinate level, these low-level features are grouped together into objects, while even further up the hierarchy these objects are grouped together as components of larger scenes involving multiple objects. When you see a red cup, for example, there will be a response on the part of neurons in your visual system that code for edges, and these neurons will represent edges at a particular location in the visual field. In addition, there will be a response on the part of neurons that code for surfaces, and there will be a response on the part of neurons that code for redness, which will represent a surface and redness at a particular area of the visual field. From one millisecond to the next, there will not be much change in these inputs, and the neural populations at the hierarchically lowest level (representing edges, surfaces and colors) may, as a default, predict no change in inputs. If the cup is moved, however, the inputs will change. Importantly, they will change in a manner that is coherent, given that they are all features of the same cup – if one of the edges moves to the left a certain distance, so will the other edges, and so will the red surface. In order to draw upon such regularities in anticipating inputs, the brain, at a hierarchical level that is superordinate to the representation of such low-level features as surfaces and edges and colors, represents the cup as an object. Moreover, to anticipate changes over longer time scales, superordinate models embed this object into larger scenes, such as tea parties, and thereby generate predictions pertaining to objects and overall scenes in a context-dependent fashion (rather than low-level features such as edges, surfaces and colors). Thus, by embedding the cup into a model of a tea party, it will become possible to predict roughly in what ways the cup will be moved, by whom, and where to. On the other hand, since we also lose detail and precision as we move up the hierarchy, lower hierarchical levels are still required in order to make specific predictions.

Modeling more abstract features of the world helps to reduce uncertainty because variance in more slowly changing causes helps to explain unexpected variance in shorter time scale causes (e.g., when the cup suddenly disappears into the dish washing machine). Cups retain their shapes for years or even centuries, as do the social norms governing behavior at tea parties. But whereas hierarchical models reduce uncertainty, prediction errors will always occur (even if one expects tea cups, for example, to be placed on tables, to be filled with tea, etc., one will generally not know *precisely* when and where). How, then, does the brain deal with the inevitable prediction errors? The basic mechanism is as follows: when a prediction error exceeds a given threshold, the model giving rise to the prediction must be revised, so an error signal is sent up to the immediately superordinate model, which is accordingly revised. New predictions are thereby generated and sent back down the hierarchy, where they are tested against new inputs. The process is repeated continuously, and in this manner the brain minimizes average long-term prediction error.[3]

When confronted with a prediction error, the brain basically has two options for reducing prediction error. The first option is to revise its model of the world until the prediction error is satisfactorily diminished. This is called 'perceptual inference'. The second option is to change the world so that it matches the model. This is called 'active inference'. If, for example, one expects to see one's cup of coffee on the desk in front of one, but it turns out not to be there, one might simply conclude that one was mistaken (i.e. change the model). But one might also adjust one's head or even one's bodily position until one does see the coffee cup, e.g. behind the laptop or occluded by a stack of books. In this case, one has changed the world in the sense of changing the position of one's body in the world. More radically, one might *go and get a cup of coffee* and put it exactly on that part of the desk where one had expected it to be. Again, this would amount to changing the world to match the model one had of it.

The concept of active inference is attractive because, together with perceptual inference, it provides a unifying framework for perception and action: both can be viewed as means of reducing prediction error. As Friston, Daunizeau, Kilner, and Kiebel (2010, p. 12) put it: "Perceptual learning and inference is necessary to induce prior expectations about how the sensorium unfolds. Action is engaged to resample the world to fulfill these expectations. This places perception and action in intimate relation and accounts for both with the same principle".

A guiding assumption of PEM is that any system that minimizes long-term prediction error will approximate Bayesian inference (Friston, 2009; see also Clark, 2013; Hohwy, 2014). In Bayesian inference, models are not only evaluated according to how well they fit the evidence (i.e., how well they predict the input in question) but also according to how likely they are in the first place (i.e., their 'prior probability'). Thus, when making sense of new sensory input, the brain does not start from scratch but, rather, updates the model with the highest prior probability in order to make it accommodate the new evidence.

3. Implications for Embodied Cognition and the Extended Mind Hypothesis

3.1. PEM and the mind-world linkage

In this section we will take a closer look at the divergent implications which Clark and Hohwy derive from PEM regarding the relation between brain, body and environment. A good starting point is the question how to balance seclusion and openness in our understanding of the relation between mind and world.

Clark (2013) recognizes that PEM offers a 'challenging vision', since it proposes that our expectations are in an important sense the primary source of what we perceive. However, he does not take this to mean that we should embrace the idea that what we perceive is the brain's best hypothesis. He claims that "what we perceive is not some internal representation or hypothesis but

[2] This is nicely illustrated in the area of reward processing by the behavior of dopaminergic neurons in the striatum: their rate of firing corresponds to unexpected changes in the value of a coming reward (e.g. increases or decreases in the number of drops of juice that are administered after a tone has sounded), not to the actual value of the reward itself (Bayer & Glimcher, 2005; Nakahara, Itoh, Kawagoe, Takikawa, & Hikosaka, 2004; Tobler, Fiorillo, & Schultz, 2005).

[3] This raises the question just how large a discrepancy between predicted and actual input can be tolerated. To deal with this issue in detail would take us too far afield, but the rough idea is that the error threshold is modulated according to the degree of expected precision. If there is a large prediction error but the signal is noisy, then there is an increased likelihood that the error is due to noise in the signal. Thus, it would be hasty to revise the model that gives rise to the prediction without further sampling. In other words, the brain engages in *second-order statistics*. This lends context-sensitivity to the system, in the hierarchical manner explained above. At twilight, for example, when conditions are not very good for vision, it is sensible to assign greater weight to one's expectations about what one is likely to encounter than on a sunny afternoon, so the threshold for prediction errors should accordingly be raised (see Hohwy, 2012 for a thorough treatment of these issues).