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

Article *in* Brain and Cognition · February 2016

DOI: 10.1016/j.bandc.2016.01.009

2 authors:

Leon De Bruin
Radboud University

John Michael
Central European University

ARTICLE IN PRESS

Brain and Cognition xxx (2016) xxx–xxx

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 xxxx

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/© 2016 Published by Elsevier Inc.

Please cite this article in press as: de Bruin, L., & Michael, J. Prediction error minimization: Implications for Embodied Cognition and the Extended Mind Hypothesis. *Brain and Cognition* (2016), http://dx.doi.org/10.1016/j.bandc.2016.01.009

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).

Please cite this article in press as: de Bruin, L., & Michael, J. Prediction error minimization: Implications for Embodied Cognition and the Extended Mind Hypothesis. *Brain and Cognition* (2016), http://dx.doi.org/10.1016/j.bandc.2016.01.009

(precisely) the world" and that "[i]t is by [inferential] means that biological beings are able to establish a truly tight mind-world linkage" (2013, p. 199). According to Clark, then, PEM introduces no worrisome barrier between mind and world.

Hohwy (2014), by contrast, argues that it is one of the central tenets of PEM that the mind is inferentially secluded from the world. In his view, PEM postulates an evidentiary boundary, a Markov blanket, which separates the worldly causes that are being inferred from the system that is doing the inferring. This results in a 'schism' between the predicting-generating models of the brain and the modeled states of affairs in the world, and makes it hard to establish the precise sense in which there can be a mind-world linkage. Because PEM ties perceptual content to a statistical model within an evidentiary boundary, it is difficult to say that what we perceive is directly or precisely the world.

This difference in interpretation, i.e. Clark's focus on openness versus Hohwy's emphasis on seclusion in our understanding of the relation between mind and world, results in a different outlook on what PEM implies for Embodied Cognition and the Extended Mind Hypothesis.

3.2. PEM and Embodied Cognition

Like proponents of moderate forms of Embodied Cognition, Hohwy is keen to emphasize the importance of the body and its environment in shaping and constraining cognition and action. He highlights two ways in which PEM accords special significance to the body. First, the concept of active inference entails a central role for the body in reducing prediction error by actively and selectively sampling the stimulus array (Friston, 2009). Second, the body must figure as a crucial parameter in the brain's model of the extra-cranial causes of its sensory inputs. This is because sensory inputs to the brain are shaped not only by causes in the world but also by the states and positions of sensory organs and bodily effectors.

However, Hohwy argues that PEM is incompatible with more radical versions of Embodied Cognition (De Jaegher, Di Paolo, & Gallagher, 2010; Gallagher, 2008; Hutto & Myin, 2013; Ratcliffe, 2007), which reject the notion of representation. This is because "The role of the body is real and substantial, but only in the sense that the body is represented in the model, as a parameter useful for minimizing prediction error" (Hohwy, 2014, p. 17).

Clark acknowledges that PEM is representational through and through. That is, he agrees with Hohwy that PEM is incompatible with radical versions of Embodied Cognition. At the same time, however, Clark maintains that (moderate versions of) Embodied Cognition can complement PEM in important ways. In particular, he points out that PEM leaves open a number of deep and important questions concerning the nature and format of human neural representation, and suggests that this is a "standing invitation to evolutionary, situated, embodied, and distributed approaches to help 'fill in the explanatory gaps'" (2013, p. 195). Addressing these questions requires a "deep (but satisfyingly natural) engagement with evolutionary, embodied, and situated approaches" (p. 200).

In other words, while Hohwy is primarily concerned with the constraints PEM places on Embodied Cognition, Clark thinks that Embodied Cognition plays an important role when it comes to the actual implementation and development of the PEM framework.

3.3. PEM and the Extended Mind Hypothesis

Hohwy argues that PEM not only presents a challenge to radical versions of Embodied Cognition but also to the Extended Mind Hypothesis. To see why, it will be useful to take a step back and recall the rationale underlying Clark and Chalmers's (1998; see also Clark, 2008) argument for the Extended Mind Hypothesis. This rests upon the following criterion for deciding whether or not an external object counts as part of a cognitive process: "if, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process" (Clark 2008, p. 76; see Adams & Aizawa, 2008 for discussion). To further explain this so-called 'parity principle', Clark and Chalmers (1998) present the following thought experiment. The fictional characters Inga and Otto both want to go to the museum. Inga remembers where it is and goes there. Otto, however, has Alzheimer's Disease and needs to consult a notebook in which he has recorded the museum's address. Clark and Chalmers suggest that there is no principled difference between the two cases: Inga's consulting her memory and Otto's consulting his notebook are both cognitive processes, since Otto's notebook is an 'external memory', literally a 'part of his mind' that resides outside his body.

According to Hohwy, PEM is incompatible with this line of reasoning because it actually does provide a principled distinction between the brain on the one hand, and the body and the environment on the other hand. Specifically, he argues that such an evidentiary boundary between the representing brain and the represented world is required for the brain to be able to perform the type of inference to the best explanation envisioned by PEM.

Hohwy gives the following example to illustrate this. Footprints in the snow outside of the house may be taken to provide evidence in favor of the hypothesis that there are burglars afoot (if this is the hypothesis with the highest posterior probability). But if further evidence indicates that the footprints were produced by pranksters, then this alternative hypothesis cannot be ruled out by appealing to the currently held belief that there are burglars afoot – because this belief is evidentially dependent upon the footprints in the snow (the status of which is currently in doubt). Appealing to the belief about burglars to rule out the hypothesis that pranksters produced the footprints would be viciously circular. In order to rule out the alternative hypothesis while avoiding a vicious circle, it is necessary to appeal to evidence that is independent of the evidence that is currently being evaluated. For the same reason, in Hohwy's view, a cognitive system must make a clear distinction between itself (as a model of the world) and the world it is modeling. This distinction is captured by the concept of a Markov blanket, which demarcates the boundary between a system engaged in inference and the evidence it draws upon.

Thus, Hohwy concludes, PEM implies an evidentiary boundary that excludes everything beyond the sensory organs: "cognitive states are not extended into the body—there is no embodied extension. Likewise, things in the environment are outside the evidentiary boundary, as are other people and their mental states. So the mind is not extended to things around us or to other people" (2014, p. 11). However, he does acknowledge that, while the existence of such an evidentiary boundary is non-negotiable, it could in principle be drawn somewhere else (such as to include the entire body as well as some external objects). The challenge to proponents of the Extended Mind Hypothesis is to define a different plausible evidentiary boundary. This boundary should make it clear that prediction error is minimized for the system under consideration, including the external object to which cognition is extended, and with respect to hidden causes outside this extended boundary. Furthermore, this minimization should happen 'on average and in the long run', because surprise minimization is defined in terms of the states the system tends to occupy in the long run.

Hohwy gives two reasons why this challenge will be difficult to meet. First, although it is possible to define the evidentiary boundary of an extended system (i.e. one that includes external objects), it is unlikely that it will be the one that best minimizes prediction error in the long run. "Whereas prediction error can be minimized

Please cite this article in press as: de Bruin, L., & Michael, J. Prediction error minimization: Implications for Embodied Cognition and the Extended Mind Hypothesis. *Brain and Cognition* (2016), http://dx.doi.org/10.1016/j.bandc.2016.01.009

transiently by systems with all sorts of objects included (e.g., shooting the tiger with a gun), on average and over the long run, it is most likely that the model providing evidence for itself is just the traditional, un-extended biological organism" (p. 12).

Second, although an external object such as a notebook can be represented by a given system, at the same time it will also be part of that system, hence "the object is both beyond one evidentiary boundary and within a further evidentiary boundary" (Hohwy, 2014). The result is that we are faced with two distinct yet overlapping agents: Otto, who represents the notebook as an external object, and Otto*, who can be defined as an extended brain-notebook system with its own evidentiary boundary. And even if we would accept the existence of multiple agents, Hohwy argues, it will still be the non-extended cognitive system (in this case, Otto's brain) that best minimizes prediction error in the long run.

If we reject the Extended Mind Hypothesis, then what do we make of the special functional role that external objects such as notebooks seem to play in our cognitive economy? Hohwy explains this in terms of 'preferential trust': the fact that the agent's expectations about these external objects are more precise than those about other objects. Preferential trust plays an important role in active inferences that involve external objects, such as those in the Inga and Otto thought experiment. For Hohwy, the principle of active inference allows us to conceptualize an intention as an expectation with high prior probability that is conditional on action, i.e. on the sensory input resulting from the causal interaction between the agent's body and the environment. Thus, Otto's intention to go to the museum can be conceptualized as an expectation about the sensory input that would result from going to the museum. However, there are different ways to get to the museum, and they come with different 'policies', i.e. different actions associated with different flows of expected sensory input. Hohwy argues that agents will typically rank their policies such that in the long run the prediction error is minimized in the most efficient way. An agent like Otto, who relies on a notebook to get to the museum, has high confidence assigned to the policy that going to the museum involves the notebook. "[T]here is preferential trust in the notebook input such that the favoured predicted sensory flow includes the sensory input arising from the agent's causal interactions with the notebook. A person with mild Alzheimer's [...] might be fully justified in assigning this policy high confidence; for example, not using the notebook will make it much harder for action to reliably fulfill the predictions of being at the museum, resulting in a prediction error increase and, in the long run, a difficulty with remaining within the expected states—such as literally getting lost" (p. 14). According to Hohwy, this way of ranking policies for active inference fully accounts for the special role certain objects play in our cognitive economy.

4. Pinpointing the sources of difference

Given that Clark and Hohwy take the same PEM framework as their starting point, it may seem strange that they arrive at such divergent assessments of Embodied Cognition and the Extended Mind Hypothesis. In this section, we will attempt to explain this by homing in on five fundamental differences between Clark and Hohwy.

4.1. The nature of representation

Although Clark seems to agree with Howhy about the representational nature of PEM, it is unclear whether they understand the notion of representation in the same way. While the concept of representation is a complex and contested one, it may be associated with the following features:

(1) Representations can be combined into a more general representational structure.
(2) Representations carry information about something other than themselves (x).
(3) Representations can misrepresent x.
(4) Representations can be decoupled from x.[4]

This notion of representation is compatible with PEM insofar as one could argue that the brain models its environment x in a way that satisfies these features. Thus, the brain can be said to (1) generate representational models that (2) attempt to predict potential input from the environment, and (3) do this in a more or less accurate way. These representational models are (4) decoupled insofar as they anticipate (represent) *potential* input from the environment.

While Hohwy does not directly address fundamental issues pertaining to the notion of representation, there is no reason to suspect that his understanding of representation differs significantly from this characterization. For Clark, however, matters are more complicated. Clark has been an advocate of so-called 'action oriented representations' or AORs – representations that are geared to drive specific sorts of actions in specific sorts of environments. AORs are context dependent in the sense that they represent knowledge of how to negotiate the environment (for a particular agent). Their vehicles do not 'stop at the skin', as Rowlands (2006) puts it, but extend 'all the way out into the world'. This, of course, fits very well with the Extended Mind Hypothesis. Yet, as we saw in the last section, it also implies that external objects that can be represented by a given system, will also be part of that very system. In representational terms, this means that they are both content *and* vehicle. Clark could try to side-step this problem by arguing that AORs are context dependent, and that the distinction between content and vehicle, between what is represented and the system that represents it, can therefore only be made when we take the relevant environment into account. Such 'situated' explanations would allow for the possibility that external objects are in some situations represented by a given system, and in other situations part of that system. However, this move puts certain restrictions on how we understand (4), i.e. the idea that a representation can be decoupled. If representations are to be individuated with respect to the environments in which they are instantiated, then they cannot be fully decoupled from these environments. This is not necessarily a problem: one could simply give up the criterion of decoupleability as part of the concept of AOR (see, for example, Wheeler, 2005). But this would result in a different notion of representation.

4.2. PEM in the long run

The core of Hohwy's argument against the Extended Mind Hypothesis consists of the claim that the brain, as a non-extended system, best minimizes prediction error in the long run. But it is precisely on this point that Clark would disagree. According to Clark, 'designer environments', i.e. human-built environments that enable new forms of reentrant processing, allow for the acquisition of generative models that do a much better job at reducing prediction error than 'their apparent base in simple forms of sensory contact with the world.' External objects play a crucial role in these environments, because they structure them in such a way so as to make them 'friendlier' for our brains. "At multiple time-scales, and using a wide variety of means (including words, equations, graphs, other agents, pictures, and all the tools of

[4] Some theories claim that representations have another feature: (5) representations are teleological, i.e. have a special function towards x (see, for example, Millikan, 1984).

Please cite this article in press as: de Bruin, L., & Michael, J. Prediction error minimization: Implications for Embodied Cognition and the Extended Mind Hypothesis. *Brain and Cognition* (2016), http://dx.doi.org/10.1016/j.bandc.2016.01.009

modern consumer electronics) we thus stack the dice so that we can more easily minimize costly prediction errors in an endlessly empowering cascade of contexts from shopping and socializing, to astronomy, philosophy, and logic" (p. 43). Thus, Clark seems to assume that extended systems (i.e. systems that include external objects) are better at minimizing prediction error in the long run than non-extended systems (i.e. the stand-alone brain). Unlike our discussion of the notion of representation in the previous section, which was mainly conceptual in nature, this assumption is in principle empirically testable.

4.3. How many agents?

Since each combination of levels within the hierarchical model can be seen as a fully formed explanatory circle with its own evidentiary boundary, PEM allows for the possibility of 'nested agents' (i.e. agents within agents). Hohwy argues that this is problematic because it leads to a proliferation of explanatory targets. He therefore proposes to rank the agents according to their overall, long-term prediction error minimization. The agent worthy of explanation is the agent that ends up at the top of this list. And this, according to Hohwy, will be the non-extended brain.

As we pointed out above, Clark would reject the assumption that the brain, as a non-extended system, best minimizes prediction error in the long run. But he may also be critical of Hohwy's explanatory monism, i.e. the claim that it is problematic to have multiple explanatory targets and the suggestion that only the agent with the highest capacity for prediction error minimization is worthy of explanation. Even if extended systems are better at reducing prediction error than non-extended systems, this does not mean that these non-extended systems are not worthy of explanation. First of all, we might be interested in the way in which a particular (non-extended) system contributes to the prediction error minimization capacity of a larger (extended) system. In Otto's case, for example, it might be important (e.g., for therapeutic purposes) to understand the specific interactions between his brain and the notebook. Second, Hohwy assumes that prediction error minimization is achieved by non-extended individual systems. Clark, however, argues that prediction error minimization can be a shared process that involves multiple agents (and not just extended individual systems): "a key proximal goal of information self-structuring, considered from the action-oriented predictive processing perspective, is the reduction of *mutual prediction error* as we collectively negotiate new and challenging domains" (2013, p. 43).

Shared prediction error minimization seems to be importantly different from (extended or non-extended) single system prediction error minimization insofar as it introduces a new level of complexity: the systems involved in shared prediction error minimization have to model each other and generate 'meta-Bayesian models', i.e. they have to make Bayesian inferences about another system's Bayesian inferences (Daunizeau, den Ouden, Pessiglione, Kiebel, Friston, et al., 2010; Daunizeau, den Ouden, Pessiglione, Kiebel, Stephan, et al., 2010; cf. also Friston & Frith, 2014). This is not only a step up in terms of computational complexity, but it also seems to increase the possibility of error between the systems involved. On the one hand, one might be inclined to argue that shared prediction error minimization allows cognitive systems to combine their resources in order to reduce mutual prediction error. This line of thought would be consistent with Clark's view that extended systems can increase their efficiency by offloading information processing and drawing upon external resources. On the other hand, however, shared prediction error minimization requires cognitive systems to deal with another source of prediction error: the existence of other cognitive systems. It is precisely for this reason that Hohwy would emphasize the additional computational burden that goes along with the dependence upon external resources (including other people). Presumably, dependence upon external resources is often sufficiently beneficial in reducing long-term prediction error to compensate for this cost.

4.4. Active inference and the desert landscape

Another difference between Hohwy and Clark has to do with their interpretation of the notion of active inference. Hohwy, following Friston (2012), notes that PEM makes it possible to assimilate desires (including goals and reward signals) to beliefs by conceptualizing them as predictions with high prior probability. Given that the brain aims to reduce prediction error in whatever way is likely to be most effective, it may go into active inference to reduce the prediction error generated when these predictions are not corroborated by incoming evidence. For example, the desire for a drink can be conceptualized as a prediction with high prior probability that is currently not corroborated by evidence. In order to reduce the resultant prediction error, the most effective option may be to go into active inference and get a drink (see Seth, 2013 for an analogous account of interoception). On this view, it is our expectations about the proprioceptive consequences of our actions (rather than our desires) that bring about our actions. Clark, on the other hand, is skeptical of what he calls the 'desert landscape' understanding of active inference. He argues that its proponents still have to demonstrate the explanatory advantages of abandoning appeals to value, reward, and cost, and of effacing the common-sense distinction between beliefs and desires. Indeed, such terms and distinctions are useful tools in the cognitive sciences, so we should not simply cast them aside without very good reasons for doing so.

4.5. Learning and experience

Given the importance accorded to expectations (priors) by any version of the PEM framework, it is crucial to explain where those expectations come from. Hohwy and Clark would both agree that the space of priors is specific to different species and varies as a result of learning and experience. Clark, though, would avow that embodied and extended perspectives are likely to be of help in identifying the structure of that space for particular species. Given that brains have evolved to make predictions that are specifically tailored to the bodies in which they reside and to the environments in which those bodies live, it may be useful and important to retain a perspective which gives pride of place to bodies and environments. Of course, for any particular instance in which bodies and/or environments play a role in an explanation of the etiology of a prior, Hohwy will also be able to accommodate it. Thus, an important challenge for Clark is to provide reasons for believing that we will be more likely to find such explanations if we endorse a perspective that systematically integrates embodied and extended resources in the explanatory framework. In particular, it would be helpful to identify case studies that demonstrate that embodied and/or extended perspectives really do bring advantages in this respect. For now, though, we can only conclude that there is an open question about just how fruitful embodied and extended perspectives will turn out to be, and that Clark and Hohwy place different bets here.

5. Outlook

While the accounts of Clark and Hohwy (unsurprisingly) overlap significantly on most central points, they also diverge markedly in the way in which they conceptualize the relation between the brain, on the one hand, and the body and its surroundings, on the other.

In this brief discussion, we have attempted to facilitate a constructive debate between proponents of 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. While we should not expect to be able to decide in a straightforward empirical manner whether the mind, as conceived be PEM, is extended or embodied, such fundamental conceptual disputes can nevertheless be fruitful in setting research agendas and in providing motivation for specific testable hypotheses. Indeed, competing answers to such fundamental conceptual questions can and should be evaluated in light of their empirical fruitfulness as well as the theoretical coherence which they make possible. Moreover, we have identified several areas in which these two rival accounts generate different empirical predictions, and shown how they can be used to formulate and structure very different research agendas, and to motivate competing hypotheses about how best to model the role of the body and/or external objects in cognition.

References

Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Malden, MA: Blackwell.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron, 47*, 129–141.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. USA: Oxford University Press.

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

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19.

Daunizeau, J., den Ouden, H. E. M., Pessiglione, M., Kiebel, S. J., Friston, K. J., & Stephan, K. E. (2010). Observing the observer (II): Deciding when to decide. *PLoS ONE, 5*, e15555. http://dx.doi.org/10.1371/journal.pone.0015555.

Daunizeau, J., den Ouden, H. E. M., Pessiglione, M., Kiebel, S. J., Stephan, K. E., & Friston, K. J. (2010). Observing the observer (I): Meta-Bayesian models of learning and decision-making. *PLoS ONE, 5*, e15554. http://dx.doi.org/10.1371/journal.pone.0015554.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences, 14*(10).

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences, 13*(7), 293–301.

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

Friston, K. (2012). A free energy principle for biological systems. *Entropy, 14*, 2100–2121.

Friston, K., Daunizeau, J., Kilner, J., & Kiebel, S. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics, 102*(3), 227–260.

Friston, K., & Frith, C. D. (2014). A duet for one. *Consciousness and Cognition*.

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition, 17*, 535–543.

Hohwy (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology, 3*, 96. http://dx.doi.org/10.3389/fpsyg.2012.00096. See also the Frontiers Research Topic, edited by N. Tsuchiya and J.v. Boxtel.

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

Hohwy (2014). The self-evidencing brain. *Nous*, 1–24. http://dx.doi.org/10.1111/nous.12062.

Hutto, D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. MIT Press.

Millikan, R. (1984). *Language, thought, and other biological categories: New foundations for realism*. MIT Press.

Nakahara, H., Itoh, H., Kawagoe, R., Takikawa, Y., & Hikosaka, O. (2004). Dopamine neurons can represent context-dependent prediction error. *Neuron, 41*, 269–280.

Ratcliffe, M. (2007). *Rethinking commonsense psychology: A critique of folk psychology, theory of mind and simulation*. Basingstoke: Palgrave Macmillan.

Rowlands, M. (2006). *Body language*. Cambridge, MA: MIT Press.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences, 17*(11), 656–663.

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science, 307*, 1642–1645.

Wheeler, M. (2005). *Reconstructing the cognitive world: The next step*. Cambridge, MA: MIT Press.

Please cite this article in press as: de Bruin, L., & Michael, J. Prediction error minimization: Implications for Embodied Cognition and the Extended Mind Hypothesis. *Brain and Cognition* (2016), http://dx.doi.org/10.1016/j.bandc.2016.01.009