Hohwy, J. (in print) The predictive processing hypothesis. (in print). In *The Oxford Handbook of 4E Cognition*. Newen, A, Bruin, L., Gallagher, S. (eds.). Oxford University Press.

The predictive processing hypothesis

Jakob Hohwy
Cognition & Philosophy Lab
Monash University
Melbourne, Australia

Abstract
Prediction may be a central concept for understanding perceptual and cognitive processing. Contemporary theoretical neuroscience formalizes the role of prediction in terms of probabilistic inference. Perception, action, attention and learning may then be unified as aspects of predictive processing in the brain. This chapter first explains the sense in which predictive processing is inferential and representational. Then follows an exploration of how the predictive processing framework relates to a series of considerations in favour of enactive, embedded, embodied and extended cognition (4e cognition). The initial impression may be that predictive processing is too representational and inferential to fit well to 4e cognition. But, in fact, predictive processing encompasses many phenomena prevalent in 4e approaches, while remaining both inferential and representational.

Introduction

A millennium ago the great arab polymath Ibn al Haytham (Alhazen) (ca. 1030; 1989), developed the view that “many visible properties are perceived by judgment and inference” (II.3.16). He knew that there are optical distortions and omissions of the image hitting the eye, which without inference would make perception as we know it impossible (Lindberg 1976, Hatfield 2002). al Haytham was aware it is counterintuitive to say perception depends on typically intellectual activities of judgment and inference and so remarks that “the shape and size of a body… and such like properties of visible objects are in most cases perceived extremely quickly, and because of this speed one is not aware of having perceived them by inference and judgment” (II.3.26).

Since al Haytham, many in optics, psychology, neuroscience, and philosophy have advocated the role of inference in perception, and have insisted too that this inference is somehow unconscious (for review, see (Hatfield 2002)). With characteristic clarity, Hermann von Helmholtz coins the phrase *unconscious perceptual inference* and says that the “psychical activities” leading to perception “are in general not conscious, but rather unconscious. In their outcomes they are like inferences insofar as we from the observed effect on our senses arrive at an idea of the cause of this effect. This is so even though we always in fact only have direct access to the events at the nerves, that is, we sense the effects, never the external objects” ((Helmholtz 1867): 430).

The starting point for this inferential view is the conviction that perception can be explained only if a particular, fundamental problem of perception is solved, namely how the brain can construct our familiar perceptual experience on the basis only of the imperfect data delivered to the senses, and without ever having unfettered access to the true hidden causes of that input. This type of problem is also at the heart of massive scientific endeavors in contemporary artificial intelligence and machine learning.

Recently, the notion of unconscious perceptual inference has been embedded in a vast probabilistic theoretical framework covering cognitive science, theoretical neurobiology, and machine learning. The basic idea is that unconscious perceptual inference is a matter of Bayesian inference, such that the brain in some manner follows Bayes' rule and thereby can overcome the problem of perception. The most comprehensive, ambitious, and fascinating of these probabilistic theories build on the notion of *prediction error minimization* (PEM) (this notion arose in machine learning research, with versions of it going back to 1950ties; for recent philosophical overviews, see (Clark 2013, Hohwy 2013)).

Several aspects of unconscious perceptual inference are anathema to many versions of enactive, embedded, embodied and extended (4e) cognition. If perception is a matter of Bayesian inference, then perception seems a very passive, intellectualist, neurocentric phenomenon of receiving sensory input and performing inferential operations on them in order to build internal representations. This process is divorced from action and active interaction with the environment, it appears insensitive to the situation in which the system is embedded, it leaves no foundational role for the body in cognitive and perceptual processes, and it makes perceptual processes a matter of what happens behind the sensory veil with no possibility of extension to mental states beyond the brain let alone the body (4e cognition is now a vast and varied area of research; the types of approaches that stress anti-representational and anti-inferential elements are, for example, (Varela, Thompson et al. 1991, Clark 1997, Noë 2004, Gallagher 2005, Thompson 2007, Clark 2008, Hutto and Myin 2013).

The tension between perceptual inference and 4e cognition matters because both are influential attempts at explaining the same range of phenomena. Having noticed the initial tension between them, there are 3 main options: (1) perceptual inference and 4e cognition are incompatible as foundational accounts of perception and cognition, which means one must be false (Anderson and Chemero 2013, Barrett 2015); this option appears unattractive because key aspects of both seem believable and important. The next two options are more discursive: (2) Perceptual inference and 4e cognition should be considered compatible, but only because perceptual inference, rightly understood, is not a matter of neurocentric, representationalist inference but yields just the kinds of processes necessary for 4e cognition (Clark 2013, Clark 2015, Clark 2016). (3) Perceptual inference and 4e cognition should be considered compatible, but only because 4e cognition, rightly understood, is nothing but representation and inference (Hohwy 2014). Options (2) and (3) deflate perceptual inference and 4e cognition, respectively, that is, they achieve reconciliation by recasting one of the sides of the debate in terms of the other.

This chapter aims to show that option (3) is reasonable. Perceptual inference, in the shape of PEM, is tremendously resourceful and can therefore encompass phenomena highlighted in debates on 4e cognition. Reconciliation with somewhat deflated 4e notions is achieved without compromising PEM's representationalist and inferentialist essence. This advances the debate about 4e cognition because, in the context of PEM, inference and representation are both shown to have several surprising aspects, such that, perhaps, 4e cognition need not abhor these notions altogether.

The chapter first explains PEM and lays out its specific notion of inference. Then action is subsumed under PEMs inferential scheme, and the role of representation in perception and action is explained. Finally, select aspects of 4e cognition are incorporated into the PEM fold.

Predictive processing and inference

On many approaches to unconscious perceptual inference, the notion of inference is left unspecified; as Helmholtz says, our psychical activities are "like" inference. Here, the notion of inference captures the idea that the perceptual and cognitive systems need to draw conclusions about the true hidden causes of sensory input vicariously, working only from the incomplete information given in the sensory input.

On modern approaches, this is given shape in terms of *Bayesian inference*. This yields a concrete sense of 'inference' where Bayes' rule is used to update internal models of the causes of the input in the light of new evidence. A Bayesian system will arrive at new probabilistically optimal "conclusions" about the hidden causes by weighting its prior expectations about the causes against the likelihood that the current evidence was caused by those causes (there are useful text book sources on machine learning, such as (Bishop 2007) and philosophical reviews (Rescorla 2015); see also recent treatments of hierarchical Bayes and volatility such as (Payzan-LeNestour and Bossaerts 2011, Mathys, Lomakina et al. 2014)).

Consider a series of sensory samples, for example auditory inputs drawn from a sound source. The question for the perceiver is where the sound source is located (somewhere on a 180° space in front of the perceiver). Assume the samples are normally distributed and that the true source is 80°. Before any samples come in, the perceiver expects – predicts – samples to be distributed around 90°. The first sample comes in indicating 77°, and thereby suggests a *prediction error* of 13°. Which probabilistic inference should the perceiver make? Inferring that the source is at 77° would disregard prior knowledge and lead to a model overfitted to noise. Ignoring the prediction error would prevent perceptual learning altogether. So the right weight to assign to the prediction error in updating the prior belief of 90° ought to reflect an optimal, rational balance between the prior and the likelihood, and this is indeed what Bayes' rule delivers. So probabilistic inference should be determined by Bayes' rule. In other words, the *learning rate* in Bayesian inference is determined by how much is already known and how much is being learned by the current evidence, reflected in the likelihood. (In this toy example, I set aside the question how the perceiver knows not to add the weighted prediction error to 90°, moving towards 103° and away from 80°; notice that if the system does this, then prediction error will tend to grow over time).

The correct weights to give to the prior and the prediction error can be considered transparently through the variance of their probability distributions. The more the variance the less the weight. A strong prior will have little variance and should be weighted highly, and a precise input, which fits well the expected values of the model in question, should be weighted highly. The inverse of the variance is called the *precision*, and it is a mathematically expedient convention to operate with precisions in discussions of inference: the learning rate in Bayesian inference therefore depends on the precisions of the priors and prediction errors. As will become apparent later, precisions are important to PEM and its ability to engage 4e type issues.

So far, only one inferential step is described. For subsequent samples, Bayes' rule should also be applied, but for the old inferred posterior as the new prior. Since there is an optimal mix of prior and likelihood, the model will converge on the true mean (80°) in the long run. Critically, in this process, the average prediction error is minimized over the long run. Even for quite noisy samples (imprecise distributions, or probability density functions), a Bayesian inference system will eventually settle on an expectation for the mean that keeps prediction error low. This can be turned around such that, subject to a number of assumptions about the shape of the probability distributions and context in which they are considered, a system that minimizes prediction error in the long run will *approximate* Bayesian inference.

The heart of PEM is then the idea that a system need not explicitly know or calculate Bayes' rule to approximate Bayesian inference. All the system needs is the ability to minimize prediction error in the long run. This is the sense in which unconscious perceptual inference is inference: internal models are refined through prediction error minimization such that Bayesian inference is approximated. The notion of inference is therefore nothing to do with propositional logic or deduction, nor with overly intellectual application of theorems of probability theory.

It would be misguided to withdraw the label 'inference' from unconscious perceptual inference, or from PEM, just because it is an approximation to Bayes, or because the process is not an explicit application of a mathematical formalism by the brain. If the inferential aspect is not kept in focus, then it would appear to be a coincidence, or somehow an optional aspect of perceptual and cognitive processes that they conform to what Bayes' rule dictate. Put differently, anyone who subscribes to the notion of predictive processing must also accept the inferential aspect. If it is thrown out, then the "prediction error minimization" part becomes a meaningless, unconstrained notion.

PEM thus says that perceivers harbor internal models that give rise to precision-weighted predictions of what the sensory input should be, and that these predictions can be compared to the actual sensory input. The ensuing prediction error guides the updates of the internal model such that prediction error in the long run is minimized and Bayesian inference approximated.

However, this description of PEM is still too sparse. In any given situation, a PEM system will not know how much or how little to weight prediction error even if it can assess the precisions of the prior and of the current prediction error. In essence, a system that operates with only those precisions will be assuming the world is more simple and persistent than it really is. For example, different sensory modalities have different precisions in different contexts, and without prior knowledge of these precisions, the system can make no informed decisions about how to weight prediction error. For example, similarly sized prediction errors in the auditory and visual modalities should not be weighted the same, since the precisions of each should be expected to be different. Therefore a PEM system would need to have and shape expectations about the precisions as well as the means of probability distributions. The need for such *expected precisions* is also driven by the occurrence of multiple interacting causes of sensory input within and across sensory modalities. In the example of the location of the auditory source, variability in the sensory sampling might be caused by a new cause interfering with the original sound source (e.g., a moving screen intermittently obscures the location of the sound). If the system does not have robust expectations for the precision of the sound source, then it will be unable to make the right inferences about the input (i.e., is it one cause with varying precisions, or is it two interacting causes that gives rise to the non-linear evolution in the auditory sensory input?).

A PEM system must model expectations of precisions, and this part of the PEM system itself needs to be Bayes optimal. Models will harbor priors for precisions, they will predict precisions and generate precision prediction errors. Moreover, it will need to do this across all the hidden causes modeled such that their interactions can be taken into account. This calls for a *hierarchical* structure, where the occurrence of various causes over many different time scales can impact on the predictions of the sensory input received at any given time. For example, the interaction of relatively slow time scale regularities (e.g., the trains driving past your house two or three times an hour) need to influence the predictions of more fast time scale regularities (e.g., the words heard in a conversation in your lounge room), and vice versa.

A PEM system that operates in a complex environment, with levels of uncertainty that depend on the current state of the world and many interacting causes at many different time scales, will thus build up a vast internal model with many interacting, hierarchically ordered levels, which all pass messages to each other in an attempt to minimize average prediction error over the long term.

Consider finally what happens over time to the models harboured in the brain, on the basis of which predictions are made and prediction errors minimized. The parameters of these models will be shaped by the Bayesian inferential process to mirror the causes of the sensory input. In the example above, by minimizing prediction error over time for the location of the cause of auditory input, the model will revise its initial, false belief that the location is at 90°, and come to expect it to be at its true position of 80°. Further, by minimizing precision prediction error, the model may be able to anticipate interacting causes, such as a moving screen intermittently blocking the sound. This means that, by approximating Bayesian inference, the models of a PEM system must *represent* its world.

Here, the notion of representation is not just a matter of receptor covariance, where the states of neural populations co-vary with the occurrence of certain environmental causes. The hierarchical model is highly structured, and performs operations over the parameters. For example, there will be model selection. In our example, the system might ask whether there is another cause interacting with the sound source, or if the signal itself is becoming more noisy. In addition, there are convolutions of separate expected signals generated on the basis of the models; for example, when a cat and a fence are detected, the expected sensory signals from both hidden causes are convolved into one stream by the brain to take the occlusion of the cat by the fence into account. As will become clear, the representational aspects PEM are critical when it comes to incorporating action too.

The representational nature of a PEM system is not optional. The ability to minimize prediction error over time depends on building better and better representations of the causes of its sensory input. This is encapsulated in the very notion of model revision in Bayesian inference. (There is extensive discussion of what it takes for perception to be representational, for examples of relevance to Bayesian inference, see (Ramsey 2007, Orlandi 2013, Orlandi 2014, Gładziejewski 2015, Ramsey 2015)).

So far, it appears that predictive processing is inferential and representational in a specific Bayesian sense. Traditionally, 4e approaches have rejected both notions. Next, PEM will be shown to have explanatory reach into 4e cognition too.

PEM and action

A representationalist and inferentialist account of cognition and perception may appear divorced from the concerns and activities of a real, embodied agent operating in its environment. Thus enactive and embodied accounts have de-emphasized classic representationalist understandings of cognition and perception and with it much semblance to inference (there are many versions and much discussion of embodiment, see, e.g., (Brooks 1991, Noë 2004, Gallagher 2005, Alsmith and Vignemont 2012, Hutto and Myin 2013, Orlandi 2014)).

Perhaps the basic sentiment could be summed up in the strong intuition that embodied action is not inference, and yet the body and its actions are crucial to gain any kind of understanding of perception and cognition. PEM can however easily cast action as a kind of inference – as *active inference* (Friston, Samothrakis et al. 2012).

Recall that any system that minimizes prediction error over time will approximate Bayesian inference, that is, such a system will be inferential in the Bayesian sense that it increases the evidence for its internal model. Using the example from above again, by minimizing prediction error the system could accumulate evidence for the model that represents the sound source as located at 80°. In that case, the internal model is revised from the initial 90° to the new estimate of 80°.

It is trivial to observe that the perceiver could also have minimized prediction error by turning the head 10° to the left and thereby have accumulated evidence for the prediction that the sound source is located at 90°. Prediction error can be minimized both through passive updating of the internal model and through active changes to the sensory input. Action, such as turning one's head, can therefore minimize prediction error. Since, as argued earlier, minimizing prediction error is inference, action is inference. There is then no hindrance to incorporating action into an inferentialist framework.

In active inference, representations are central to guiding action. This is because action only occurs when a hypothesis – in this case a representation of a state that is yet to occur – has accumulated sufficient evidence relative to other hypotheses to become the target of prediction error minimization. This yields two aspects that are sometimes seen as hallmarks of representations: they are action guiding and they are somehow detached from what they stand for (for discussion and review, see (Orlandi 2014)). Active inference therefore has a good claim to be both inferential and representational.

For perceptual inference, precisions were shown to be critical. Without precisions, the PEM system would not be able to minimize error in a world with state-dependent uncertainty and interacting causes. The same holds for active inference. Without any notion of how levels of prediction error tend to shift over many interacting time scales, the system would pick the action that minimizes most error here and now – for example by entering and remaining in a dark room (for discussion, see (Friston, Thornton et al. 2012)). This would be analogous to overfitting, and would come at the cost of increasing prediction error over the longer term. For example, even though the perceiver might minimize prediction error by forcing the sound to come at the 90° midline, this might make it difficult to ascertain the true source of a potentially moving cause such as the trajectory of a mosquito buzzing about (since direction detection is harder over the midline due to minimal interaural time difference). This calls for even more hierarchical model building, namely in terms of the precisions expected in the evolution of the prediction error landscape as a result of the agent's active intervention in the world. These self-involving, modeled regularities are however not fundamentally different to the regularities involved in perceptual inference. They simply concern the sensory input the agent should expect to result from the interaction of one particular cause in the world – the agent itself – with all the other causes of sensory input (for discussion of self-models see, e.g., (Synofzik, Vosgerau et al. 2008, Metzinger 2009)).

There is thus room for a notion of action within PEM. But this possibility alone does not imply that a PEM system is likely to actually *be* an agent. If the system is endowed with a body such that it could act, then the imperative for minimization of prediction error will make actual action highly likely.

If the system has accumulated strong evidence for, say, an association between two sounds, it may still be unable to distinguish several hypotheses, for example whether the sounds are related as cause and effect or if they are effects of some common cause. It is standard in the causal inference literature that intervention is required to acquire evidence for or against these hypotheses (Pearl 2000, Woodward 2003). For example, if variation in one sound persists even if the other sound is actively switched off, then that is evidence the latter sound is not the cause of the first. The necessity of action is generalized in the observation from earlier that the system needs to learn differences in precisions and patterns of interactions amongst causes, such as occlusions and other causal relations that change the sensory input in nonlinear ways. Such learning thus requires action. The prize of not engaging the body plant to intervene in the environment is that prediction error will tend to increase since predictions will be unable to distinguish between several different hypotheses. A PEM system that can act will therefore be best served to actually act.

This simple account of agency has profound consequences. It will be a learnable pattern in nature that inaction will tend to increase prediction error in the longer term (due to the inaccuracy of the hypotheses the system can accumulate evidence for by using only passive inference). Conversely, the system can learn that action tends to allow minimization of prediction error at reasonable time scales. Overall, this teaches the system that, on balance, its model will accumulate more precise evidence through action than through inaction. This will bias it to minimize prediction error through active inference. Of course, a system that only ever acts on the basis of unchanging models will never be able to learn new patterns, which is detrimental in a changing world. Therefore action must be interspersed with perceptual inference where models are updated, before new action takes place.

The mechanism by which this switching between perception and action takes place is best conceived in terms of precision optimization. Recall that the PEM system will build up expectations for precisions, which are crucial for dealing with state-dependent noise in a world with interacting causes. The role of expected precisions in inference is to optimally adjust weights for expected sensory input: input that is expected to be precise is favoured in Bayesian inference whereas input that is expected to be imprecise is not favoured. Mechanistically, this calls for a neuronal gating mechanism that inhibits or excites sensory input according to their expected precisions. This gating mechanism serves as a kind of probabilistic search-light and thus plays the functional role of *attention* (Feldman and Friston 2010, Brown, Friston et al. 2011, Hohwy 2012, Hohwy 2016).

As the system gates its sensory input according to where it expects the most precise sensory input will occur, across several time scales, it may switch between perception and action. For example, if more precision is expected by the agent having its hand at the position of the coffee cup rather than at the current position at the laptop, then it will begin gating the current sensory input, which suggests the hand is at the laptop. This in turn allows the coffee-hypothesis to gain relative weight over the laptop-hypothesis, and the prediction error generated by that hypothesis can easily by minimized by moving the hand. Since the gain is high on this prediction error, the new hypothesis quickly accumulates evidence for its truth, and the hand will find itself at the coffee cup (for more on the dynamics of action and perception in relation to temporal phenomenology, see (Hohwy, Paton et al. 2015), for the formal background, see (Friston, Trujillo-Barreto et al. 2008)).

Embodied, embedded, and inferential and representational.

When all the elements described in the last section are combined, a wholly inferential conception of agency begins to take shape. If action and agency are moments of prediction error minimization, then desires are just beliefs (or priors) about states that happen to be future, with a focus on their anticipated levels of prediction error, and where reward is the absence of prediction error. This suggests a neat continuity with perceptual inference, which also relies on priors and the imperative to minimize prediction error.

The idea that action is driven by prediction error minimization relative to a model does raise a question about the content of the model relative to which error is minimized. This model is what defines what we would normally describe as the agent's desires. In the wider PEM framework, which, as shall be described below, relies on notions of *free energy minimization*, the expected states that anchor active inference relate to set points in terms of the organism's homeostasis. This immediately evokes an evolutionary perspective, where expected bodily states are central to behavior. Apart from the specific evolutionary aspects, this suggests an *embodiment* perspective because all aspects of perception and cognition then have a foundation in bodily states, and movement and purposeful behaviour in the environment. This element of embodiment makes it more likely that contact can be made between probabilistic theories of perception and action and embodied cognition approaches (such as, e.g., (Varela, Thompson et al. 1991, Gallagher 2005, Thompson 2007); for recent treatments that relate to PEM, see (Bruineberg and Rietveld 2014, Fazelpour and Thompson 2015)).

However, even this foundational embodiment is conceived probabilistically in PEM. A set of expectations for bodily states (relating to homeostasis) is essentially a model. In probabilistic terms, this model gives the probability of finding the organism in some subset of the overall set of states it could be in. The model is specified in terms of internal states, as signaled in interoception, but is tied to the overall setting of the organism in a subset of environmental states. The expected states defined in interoceptive terms would, in real organisms traversing actual environments, be mirrored in the expected states described in environmental terms, or in terms of their sensory input or exteroception. For example, fish are most likely find their sensory organs impinged upon from watery states and this is associated strongly with the homeostatic needs specified in their model. In general, within this probabilistic reading of the foundational embodiment of a PEM organism, there is thus a tight coupling between the interoceptive and exteroceptive prediction error landscapes for any PEM system.

Not only does PEM provide a notion of embodiment, it also speaks to elements of embedded or situated cognition (see (van Gelder 1995, Clark 1997, Aydede and Robbins 2009)). With the tight coupling of the organism's expected states in terms of interoception and exteroception, perception and cognition cannot be separated from bodily nor environmental aspects of the PEM system.

Crucially, this reading of embodiment and embedding leads directly to inferential processing and PEM. The model specifies the probability of finding the organism in any one of all the possible states. To know this model directly would require the agent averaging over all possible states and ascertain the occurrence of itself in them. This is not possible for a finite organism to learn directly. Instead, the organism must essentially guess what its expected states are and minimize the ensuing error through perceptual and active inference. In slightly more formal terms, the organism needs to minimize surprise, that is, it needs to avoid finding itself in states that are surprising given its model. The sum of prediction error is always equal to or larger than the surprise, so minimizing prediction error will implicitly minimize surprise. This bound on surprise is also known in probabilistic terms as the free energy, and so this challenging idea is enshrined in the so-called *free energy principle* (Friston 2010).

When viewed in this larger context of the free energy principle, promising notions of embodied and embedded cognition present themselves. More research is needed on the extent to which they capture facets of the wide-ranging and heterogeneous 4e body of research. However, for the conception of embodiment and embedding mooted here, an inferential conception is inescapable.

Hierarchical inference for a changing world

In much 4e research there is a focus on fluid interactions with the world, characterized by non-inferential, non-representational, "quick and dirty" processing. This picture is set up to contrast with inferential, representational, "slow and clean" processing (Clark 1997, Clark 2013, Clark 2015). Often, this kind of quick and dirty, situated cognition is discussed in terms of *affordances*: salient elements of the environment that are in some sense perceived directly and are immediately action guiding. Affordances in quick and dirty processing are thought to evade the computational "bottleneck" that a traditional representational system would have trying to passively encode the entire sensory input presented at any given time. For some types of action and at some stages of learning, performance is rather plodding and sluggish, but there is an important insight in how the notion of situated cognition highlights the fluid swiftness with which organisms can perform some complex actions in their environment.

In a PEM system there is no bottleneck problem in the first place, however. There is never an issue of starting from scratch and encoding an entire natural scene in order to be able to perceive it. Hierarchical Bayesian inference is based on prior learning, which over time has shaped priors at many levels. Given priors, the sensory input is no longer something that needs to be encoded here and now. Instead the sensory input is, functionally speaking, the *feedback* to the forwards predictive signal generated by the brain's internal model (Friston 2005). The model predicts what will happen and gets confirmation or disconfirmation on these predictions from the sensory input. There is thus no encoding of the entire sensory input in each perceptual instance. This means the PEM system has no need to resort to quick and dirty processing tricks to overcome a computational bottleneck. Instead, the system relies on slow and clean learning in order to facilitate swift and fluid perception and interaction with the world. This learning is 'slow' because is relies on meticulous accumulation of evidence for hypotheses at multiple time scales. It is 'clean' because the learning slots into a hierarchy with clearly defined, general functional roles for time scales, for predictions of values, and for predictions of precisions.

The difference between swift and fluid processing and plodding and sluggish processing can easily be accommodated within a PEM system. Affordances are just causes of sensory input that, on the basis of prior learning, are strongly expected to give rise to high precision prediction error. To maintain Bayes optimality, this gates sensory input accordingly, and strongly focuses both perceptual and active inference on these affordances. In this setting, prediction error minimization happens quickly, since highly precise distributions are easier to deal with computationally than imprecise ones. This means that the agent in question will obtain its expected states swiftly and fluidly.

Typically, the 4e preference for quick and dirty processing and affordances comes with a rejection of rich representational states (Clark 2008, Clark 2015). The point is that such representations cannot come about due to the bottleneck problem. Moreover, the appeal to affordance-based quick and dirty processing is thought to obviate the need for rich internal representations altogether as the world's affordances in some sense is its own representation (Brooks 1991).

On the PEM-based account of swift and fluid processing, internal representations are however necessary. Over time, multi-layered representations are constructed and shaped, and Bayesian model selection pick the model with the best evidence as the representation of the world relative to which prediction error is minimized in active inference (this kind of approach is developed in more detail for PEM in (Seth 2014, Seth 2015)). Again, we get the result that PEM has the resources to speak to typical 4e discussions, but that it happens on the basis of representation and inference.

It could be that the brain builds rich representations as it learns about the world, and then gradually substitutes these much more sparse and representation poor, purpose-made representations that more directly tie in with and engage the environment. One argument here derives from Occam's razor, in the sense that there are simplicity gains from opting for a simple over a complex, rich model (Clark 2015). However, simplicity is not something additional to inference. Complex models are to be avoided because they are overfitted and thereby incur a prediction error cost in the longer run. How rich or simple a model should be is thus fully given by PEM in the first place.

In fact, there is reason to think the PEM account is preferable to the affordance-based account. It is true that swift and fluid processing is a salient and impressive aspect of human cognition. But so is the flexible way we shift between contexts, projects, beliefs, and actions. We might engage in attentive, fluid and swift interaction for a period of time but other beliefs and concerns always creep in and make it imperative to shift to another behavior. On the affordance-based account it is not readily explained how the agent might disengage from a given set of affordances; the focus is at best on how representation rich learning is needed before swift and fluid processing is possible, rather than the role of rich representation during swift and fluid processing. The agent seems tightly knitted to its environment, and it is not clear how the agent can step back and reconsider its current course of action.

In contrast, flexible cognition is a central motivation for adopting PEM's hierarchical Bayesian inference in the first place. Active inference is driven by the most probable hypothesis at any given time. The system will have built up expectations not just for what the most likely causes of sensory input might be but also for the typical evolution of prediction error precision. In particular, there will be accumulated evidence that any given hypothesis under which prediction error is minimized at a certain time will have a limited life span – in essence the system will know that it lives in a changing world where precise evidence for any given hypothesis will soon begin to be hard to find. For example, as the agent fluidly and swiftly catches baseballs it will know that the sun will soon set and make the visual input imprecise. It will therefore begin accumulating evidence for the next hypothesis (e.g., "I am eating dinner") under which evidence will soon begin to be accumulated and prediction error minimized.

This speaks to a crucial balance, which a PEM system must obtain. As prediction error is minimized in active inference, the hypothesis relative to which error is minimized is held stable. This means that, as prediction error is minimized, the world can in fact change "behind the scenes" to such an extent that it would eventually be better to abandon the current hypothesis and adopt a new one. Anticipating such change in the environment matters greatly to the agent because it should never engage in any behavior, no matter how swift and fluid, for so long that when it ceases the behavior, the world has changed in other respects and predictive error will be very large. A PEM agent therefore will be inclined to believe that the current state of affairs will change, and therefore the agent will intersperse active inference with perceptual inference, where the internal model is checked and the size of the overall prediction error is adjusted and tightened up before a new hypothesis is selected for active inference (see (Hohwy 2013, Hohwy, Paton et al. 2015).

A hierarchical system operating with slow and clean processing can thus economically explain both swift and fluid, affordance-based cognition as well as flexible cognition. This is an important point to make in the context of PEM's affinity to 4e cognition. The motivation for PEM is, in the end, the simple observation that we live in a changing world. Our world presents many different causes of our sensory input, and these causes interact with each other to create non-linearities in the input; moreover, these interactions happen concurrently at many different time scales (e.g., "The setting sun makes the balls hard to see but this time of the year the janitor often turns on the flood lights at the far pitch…"). This complexity is what creates the need for hierarchical Bayesian inference in the first place: a rich internal model that keeps track of all these contingencies and can mix the various causes in the right way to anticipate the sensory input. This has a 4e-type ring to it: the cognitive system is the way it is because the agent's world and body is the way it is. In particular, PEM is not the best solution for non-ecological, lab-style model environments where typically context and interactions between hidden causes is kept to a minimum. In other words, a machine learning researcher who never test their system against the real world will have little impetus to build a PEM system. On 4e approaches, there is also a strong focus on real-world settings but the response is typically to tie the agent very closely to its environment. This however makes it harder to see how not just the real world, but also that fact that the real world is a changing place, can be taken into consideration. PEM, in contrast, makes room for the changing world by retracting further away from the world, into a vast internal model that seeks to represent the full richness of the world and the way it changes over many time scales. On the PEM conception of the agent's place in the world, cognition is not a matter of being closely in tune with and driven by the sensory input. Rather, cognition is a matter of having richly represented expectations for the world and the body and seeking confirming feedback on those expectations through the senses.

The mind and things without it

Both perception and action are inferential and representational. The PEM system's process of minimizing prediction error implies that the sensory input is explained away on the basis of the evolving hypotheses of an internal model. The more the system can minimize its prediction error, the more it will accumulate evidence for its own truth. This is a trivial observation: if I can minimize prediction error for my theory that my hamster has escaped the more evidence I have for that theory. If we consider the PEM system an agent, then it acquires evidence for its own existence through its activities (Friston 2010). Borrowing a term from philosophy of science, the PEM system can thus be said to be *self-evidencing* (Hempel 1965, Hohwy 2014).

A self-evidencing system creates a sensory boundary between itself (i.e., the model) and the causes of its sensory input. This again is a trivial consequence of self-evidencing: there is something that garners evidence and then there is what the evidence is evidence of. Or again, in both perceptual and active inference there is something doing the inference and something being inferred. This boundary can also be described in terms of causal nets, where a set of inner states (i.e., brain states) can be said to have a *Markov blanket* (Pearl 1988) consisting of the inner states' parents (i.e., the sensory states), and their children and other parents of the children (i.e., the active states driving active inference) ((Friston 2013, Hohwy 2015); causal Bayes nets must be acyclic but brains have recurrent (cyclic) states; there are technical ways, such as dynamical Bayes nets deal with such problems). The activity of the states within a Markov blanket is wholly determined the states of the blanket. In principle, nothing about the environmental states beyond the blanket need be known to know what the system is doing. By extension, in principle, only the states of the sensory organs need be known to know everything the mind does.

PEM then comes with a principled way of drawing a boundary between the mind and the outside world. If a particular state is part of what is doing the inference, then it must be within the sensory boundary, as a part of what approximates inference about outside causes of sensory input. This may relate to the vigorous debate about *extended cognition* (Clark and Chalmers 1998, Clark 2008), which is the last member of 4e cognition to discuss.

Extended cognition is the idea that some objects, such as notebooks and smart phones, play such an integrated, memory-like function in the mental economy of some agents that, by parity of reasoning, they should be considered part of the agent's mental states even though they reside outside the central nervous system. There is much discussion of this idea (see, e.g., (Menary 2007, Adams and Aizawa 2008, Anderson, Richardson et al. 2012, Spaulding 2012)). PEM brings with it a new way of thinking about the role of such external objects. On the one hand, these objects are inferred (e.g., on the basis of the sensory input from the notebook) and as such they are outside the mental states of the system. On the other hand, if the extended cognition hypothesis is correct, they are within the sensory boundary, forming part of the inner states behind a Markov blanket inferring the hidden causes beyond it.

Interpreting purported cases of extended cognition according to PEM thus leaves two main options. There might be contradiction, since something cannot be both within and beyond the same boundary at the same time. Or, there might be multiple co-existing sensory boundaries. The second option is very interesting and very likely to be true, since Markov blankets occur easily. There is an associated cost however: we have identified the inner states (or the model) with the agent, and if there are multiple Markov blankets then there are multiple agents co-existing at the same time. Though this may be true in a weak sense of agent, it is explanatorily messy. When asking which agent is acting, there would then be a multitude of correct answers, depending on how many nested Markov blankets are involved in the same action. This speaks in favour of using inference to the best explanation to identify the agent whose relatively invariant involvement accounts for most of observed behavior over time. It seems likely this more pragmatically identified agent would be the agent as specified by the model harboured just in the nervous system. This is the agent relative to which prediction error is minimized over the longer time scale, which as we saw is central to understanding predictive processing accounts in the first place (for discussion, see (Hohwy 2014)). Bringing this discussion back to extended cognition, the pragmatic method of identifying the agent suggests that there is no extended cognition, since the special objects in question are beyond the one Markov blanket. The more lax way of identifying agents suggests that extended cognition ambiguous, since the special objects are beyond some blankets and within others.

The existence of the sensory boundary or Markov blanket implies that perception and agency are confined to the inner states of the PEM system (wherever the boundary or boundaries of the system is located). Those inner states will mirror the states outside the boundary: the inner states will, through prediction error minimization, come to represent the worldly causes of the sensory input impinging at the system's periphery. Conversely, through active inference, the outside states will come to conform to the expectations harboured in the internal states.

There is then an intriguing duality to this sensory boundary between mind and world. On the one hand, the boundary is epistemic (cf. self-evidencing): the worldly causes can only be known vicariously, through inference on sensory input. On the other hand, the boundary is characterized in causal terms (cf. Markov blanket): there is a dynamic coupling between mind and world, enabled through both perception and action.

This duality summarises well why PEM is a good fit for many of the issues in 4e debates: PEM is able to throw light on embodied agents dynamically interacting with the environment in which they are embedded. This good fit with 4e cognition is however made possible precisely because PEM is inferential and representational.

Adams, F. and K. Aizawa (2008). The Bounds of Cognition. Oxford, Blackwell.
Alsmith, A. J. T. and F. Vignemont (2012). "Embodying the Mind and Representing the Body." Review of Philosophy and Psychology **3**(1): 1-13.
Anderson, M. and A. Chemero (2013). "The problem with brain GUTs: Conflation of different senses of "prediction" threatens metaphysical disaster." Behavioral & Brain Sciences **36**: 204-205.
Anderson, M. L., M. J. Richardson and A. Chemero (2012). "Eroding the Boundaries of Cognition: Implications of Embodiment." Topics in Cognitive Science: 717–730.

Aydede, M. and P. Robbins (2009). The Cambridge handbook of situated cognition. New York, Cambridge University Press
Barrett, L. (2015). "A Better Kind of Continuity." The Southern Journal of Philosophy **53**: 28-49.
Bishop, C. M. (2007). Pattern Recognition and Machine Learning. Cordrecht, Springer.
Brooks, R. A. (1991). "Intelligence without representation." Artificial Intelligence **47**(1–3): 139-159.
Brown, H., K. J. Friston and S. Bestmann (2011). "Active inference, attention and motor preparation." Frontiers in Psychology **2**.
Bruineberg, J. and E. Rietveld (2014). "Self-organization, free energy minimization, and optimal grip on a field of affordances." Frontiers in Human Neuroscience **8**.
Clark, A. (1997). Being There. Cambridge, Mass., MIT Press.
Clark, A. (2008). Supersizing the Mind: Embodiment, Action, and Cognitive Extension, Oxford University Press, USA.
Clark, A. (2013). "Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science." Behavioral & Brain Sciences **36**(3): 181-204.
Clark, A. (2015). "Radical Predictive Processing." The Southern Journal of Philosophy **53**: 3-27.
Clark, A. (2016). Surfing Uncertainty. New York, Oxford University Press.
Clark, A. and D. Chalmers (1998). "The Extended Mind." Analysis **58**(1): 7-19.
Fazelpour, S. and E. Thompson (2015). "The Kantian brain: brain dynamics from a neurophenomenological perspective." Current Opinion in Neurobiology **31**(0): 223-229.
Feldman, H. and K. Friston (2010). "Attention, uncertainty and free-energy." Frontiers in Human Neuroscience **4**(215).
Friston, K. (2010). "The free-energy principle: a unified brain theory?" Nat Rev Neurosci **11**(2): 127-138.
Friston, K. (2013). "Life as we know it." Journal of The Royal Society Interface **10**(86).
Friston, K., S. Samothrakis and R. Montague (2012). "Active inference and agency: optimal control without cost functions." Biological Cybernetics **106**(8): 523-541.
Friston, K., C. Thornton and A. Clark (2012). "Free-energy Minimization and The Dark Room Problem." Frontiers in Psychology **3**.
Friston, K. J. (2005). "A theory of cortical responses." Philosophical Transactions: Biological Sciences **369**(1456): 815 - 836.
Friston, K. J., N. Trujillo-Barreto and J. Daunizeau (2008). "DEM: A variational treatment of dynamic systems." NeuroImage **41**(3): 849-885.
Gallagher, S. (2005). How the Body Shapes the Mind. Oxford, Oxford University Press.
Gładziejewski, P. (2015). "Predictive coding and representationalism." Synthese: 1-24.
Hatfield, G. (2002). Perception as unconscious inference. Perception and the physical world. D. Heyer and R. Mausfeld, John Wiley & Sons, Ltd.**:** 113-143.
Helmholtz, H. v. (1867). Handbuch der Physiologishen Optik. Leipzig, Leopold Voss.

Hempel, C. G. (1965). Aspects of Scientific Explanation and Other Essays in the Philosophy of Science. New York, Free Press.
Hohwy, J. (2012). "Attention and conscious perception in the hypothesis testing brain." Frontiers in Psychology **3**.
Hohwy, J. (2013). The Predictive Mind. Oxford, Oxford University Press.
Hohwy, J. (2014). "The Self-Evidencing Brain." Noûs: n/a-n/a.
Hohwy, J. (2015). The neural organ explains the mind. Open MIND. T. Metzinger and J. M. Windt. Frankfurt am Main, MIND Group**:** 1-23.
Hohwy, J. (2016). Prediction, agency, and body ownership. Where is the action? The pragmatic turn in cognitive science. K. Friston. Cambridge, Mass., MIT Press.
Hohwy, J., B. Paton and C. Palmer (2015). "Distrusting the present." Phenomenology and the Cognitive Sciences: 1-21.
Hutto, D. and E. Myin (2013). Radicalizing enactivism: Basic minds without content. Cambridge, Mass., MIT Press.
Mathys, C. D., E. I. Lomakina, J. Daunizeau, S. Iglesias, K. H. Brodersen, K. J. Friston and K. E. Stephan (2014). "Uncertainty in perception and the Hierarchical Gaussian Filter." Frontiers in Human Neuroscience **8**.
Menary, R. (2007). Cognitive integration: Mind and cognition unbounded. Basingstoke, Palgrave Macmillan.
Metzinger, T. (2009). The Ego Tunnel. New York, Basic Books.
Noë, A. (2004). Action in Perception. Cambridge, Mass., MIT Press.
Orlandi, N. (2013). "Embedded Seeing: Vision in the Natural World." Noûs **47**(4): 727-747.
Orlandi, N. (2014). The innocent eye: Why vision is not a cognitive process. Oxford, Oxford University Press.
Payzan-LeNestour, E. and P. Bossaerts (2011). "Risk, Unexpected Uncertainty, and Estimation Uncertainty: Bayesian Learning in Unstable Settings." PLoS Comput Biol **7**(1): e1001048.
Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San fransisco, Morgan Kaufmann Publishers.
Pearl, J. (2000). Causality. Cambridge, Cambridge University Press.
Ramsey, W. (2007). Representation reconsidered. Cambridge, Cambridge University Press.
Ramsey, W. (2015). "Must cognition be representational?" Synthese: 1-18.
Rescorla, M. (2015). Bayesian Perceptual Psychology. Oxford Handbook of the Philosophy of Perception. M. Matthen. Oxford, Oxford University Press**:** 694-716.
Seth, A. K. (2014). "A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia." Cognitive Neuroscience **5**(2): 97-118.
Seth, A. K. (2015). The Cybernetic Bayesian Brain. Open MIND. T. K. Metzinger and J. M. Windt. Frankfurt am Main, MIND Group.
Spaulding, S. (2012). "Overextended cognition." Philosophical Psychology **25**(4): 469-490.
Synofzik, M., G. Vosgerau and A. Newen (2008). "I move, therefore I am: A new theoretical framework to investigate agency and ownership." Consciousness and Cognition **17**(2): 411-424.
Thompson, E. (2007). Mind in Life: Biology, Phenomenology, and the Sciences of Mind. Harvard, Harvard University Press.

van Gelder, T. (1995). "What might cognition be, if not computation?" Journal of Philosophy **91**: 345–381.
Varela, F., E. Thompson and E. Rosch (1991). The Embodied Mind. Cambridge, MA., MIT Press.
Woodward, J. (2003). Making Things Happen. New York, Oxford University Press.