Busting Out: Predictive Brains, Embodied Minds, and the Puzzle of the Evidentiary Veil

Andy Clark
School of Philosophy, Psychology, and Language Sciences
University of Edinburgh, UK

Abstract

Biological brains are increasingly cast as 'prediction machines': evolved organs whose core operating principle is to learn about the world by trying to predict their own patterns of sensory stimulation. This, some argue, should lead us to embrace a brain-bound 'neurocentric' vision of the mind. The mind, such views suggest, consists entirely in the skull-bound activity of the predictive brain. In this paper I reject the inference from predictive brains to skull-bound minds. Predictive brains, I hope to show, can be apt participants in larger cognitive circuits. The path is thus cleared for a new synthesis in which predictive brains act as entry-points for 'extended minds', and embodiment and action contribute constitutively to knowing contact with the world.

1. Predictive Processing

The vision of the brain as a biological engine of prediction is steadily gaining ground (Bubic, von Cramon, & Schubotz, 2010; Andy Clark, 2013; K. Friston, 2005, 2010; Hohwy, 2013). Brains, this emerging vision suggests, are fundamentally prediction-error minimizing devices. Prediction error, in these treatments, signals the mismatch between incoming sensory stimulation and complex, multi-area downward flows of neuronal activity. Brains like these are complex self-organizing systems that alter and change so as actively to predict the incoming sensory barrage. This is not, for the most part, a matter of trying to look into the future! Instead, such systems are constantly trying to *guess the present*: they are trying to self-generate ('from the top down') the sensory streams that are currently arriving from the world. Call this general strategy 'engaging in the prediction task'.

The prediction task is computationally important because it opens the door to powerful forms of unsupervised learning. A system that initially knows nothing (or very little[1]) about its world can still progressively alter its own structure so as to meet incoming sensory signals with increasingly apt flows of top-down prediction. It does this by automatically altering so as to reduce future mismatches between input and the top-down flow. This process (which can be implemented by various well-understood algorithms[2]) slowly installs a kind of statistical model of the most likely ways in which sensory inputs will change and evolve. Such models are often tracking distal causes. If you see (sense) the flash of lightning and have learnt to expect the distinctive auditory sensory barrage caused by thunder, your statistical model works because it is tracking a regularity in the distal environment. If you hear a noun and expect a verb, your statistical model works because it is tracking a regularity in the linguistic environment. In a variety of simulation studies it has been shown that self-organizing around the prediction task enables systems that start with little or no grip on such regularities to gain such a grip[3].

Such systems learn about the shape of the distal environment by using some of their states (the flows of downward prediction) to predict the evolving shape of others (ultimately, states of their own sensory surfaces). When such guessing games are conducted systematically, using multi-level architectures in which each level is trying to guess (predict) the activity at the level below, the results are impressive[4]. Each higher level learns to predict the activity at the level below by learning about different kinds of structure in that activity – structure that, layer-by-layer, reaches back to the structure of perturbations at the sensory surfaces. In this way, they come to make increasingly 'educated guesses' about the world - the distal structures that are the source of those patterned input signals. Exposed only to raw sensory stimulations, hierarchical predictive learning regimes thus deliver a grip upon complex structures of interacting distal causes (structures of 'hidden causes' or 'latent variables') as the most efficient means of predicting future plays of energy across the sensorium.

Potent regimes of prediction-driven learning, these stories suggest, enable brains like ours to learn about what's out there by finding the best ways to predict the incoming sensory barrage using linked bodies of acquired probabilistic knowledge. Linked bodies of probabilistic knowledge capable of performing this feat are known as 'generative models' (see e.g (Hinton, 2007)) because they are able to re-create the incoming sensory patterns 'from the top-down'. Exposed to certain streams of auditory information, such systems might learn about words, phrases, nouns, verbs, and sentences. Exposed to certain streams of visual information, they might learn about basic and more complex structures found in natural scenes (lines, edges, stripes, and ultimately whole objects). Exposed to certain streams of proprioceptive and interoceptive information, they might (Seth, 2013) learn about the shapes, structures, and capabilities of their own bodies. Prediction-driven learning thus provides a mechanistically tractable means of learning about the distal world on the basis of nothing more than the shifting patterns of raw energies registered by the sensory transducers. When a well-trained system later encounters (or brings forth –see below) a pattern of sensory stimulation, it tries to match that pattern using what it has learnt to generate a flow of top-down prediction. Mismatches between what is predicted and what is currently registered at the sensory surfaces yield prediction errors, and those prediction errors are used to nuance the ongoing process of multi-level top-down guessing until an acceptable match (corresponding to a reasonable percept) is achieved. In such 'predictive processing' (PP), or 'prediction error minimizing (PEM[5]), regimes incoming sensory signals are met using a rich suite of knowledge and expectation whose content and deployment is controlled by (different time-scale) processes of prediction-error minimization. These processes minimize the difference between what is predicted and the sensory input as it is registered by (ultimately) the plays of energy across the sensory surfaces (for introductions, see (Andy Clark, 2013; Hohwy, 2013)).

The PP (Predictive Processing) story is powerful and progressive. It accounts for a wide variety of observed effects and reveals complex inter-relations between core cognitive phenomena such as attention, learning, perception, action, and cognition[6]. It appears increasingly likely that PP routines pervade human cognition, and that it is the systematic pursuit of apt top-down prediction that installs, in brains like ours, richly structured suites of knowledge concerning the wider world. If these ambitious accounts are on track the ongoing attempt to meet the incoming sensory stream with matching patterns of top-down activation provides both the core mechanism responsible for bringing a structured external reality 'into view' in the first place and the means by which ongoing perceptual experience is itself constructed.

2. The Bounds of Sense

Prediction error minimization takes place behind what (Hohwy, 2013, 2014) terms an 'evidentiary boundary'. In one sense, this is clearly true. Prediction error is minimized, ultimately, for the plays of stimulation that occur (or are actively brought forth) at the sensory surfaces, wherever they may be. That marks a boundary – though it need not mark a boundary that is unique or immutable, as we shall later see. In addition, Howhy (e.g. (Hohwy, 2013) p.239) speaks of these systems as seeking the 'hypotheses' that best explain the sensory data. This too is true – but only (as I hope to show) on a relatively weak reading of the term 'hypothesis'. This will prove important as we seek to uncover the deeper implications of the PP/PEM story.

The predicting brain seems to be in somewhat the same predicament as the imprisoned agents in Plato's "allegory of the cave" (Plato. & Jowett, 1941). Plato imagines agents chained to a wall, able to see only the shadows cast as events unfold in front of a flickering fire. The imprisoned cave-dwellers are thus unable to directly perceive events in the world, and must instead guess at the world on the basis of the distorted shadows. Similarly, Hohwy argues, the biological brain/central nervous system must constantly infer the shape and structure of the distal realm on the basis of the partial and fragmentary information available in the sensory signal.

The cave-dwellers are, notably, prisoners in the cave – they are chained to the walls, unable to intervene in the scenes they are watching. Human agents, by contrast, can act upon the world in ways that test their hypotheses – for example, we may saccade to the locations where we expect to find salient visual information given that we think we are engaging in face-to-face conversation. If we find the kind of information we expect (e.g. flows of information from the eye and face regions indicating changing emotional states) that supports the 'face-to-face conversation' hypothesis. If we do not, we may have to seek a new hypothesis (perhaps I mistook a showroom dummy for a real agent). By actively interrogating the world, we thus put our hypotheses to the test[7].

It might be thought that this makes all the difference. But as (Hohwy, 2014) also notes, the mere availability of action does not materially alter the epistemic situation for the predictive brain. For action itself, according to the PP class of models, is best understood as a means of minimizing prediction error. In action, prediction error is minimized not by altering our hypotheses about the world, but by altering the world so as to make it conform to our predictions. The idea here (see (K. J. Friston, Daunizeau, Kilner, & Kiebel, 2010; K. Friston, 2009) is that action is brought about by proprioceptive prediction. The brain predicts the flow of proprioceptive consequences that would obtain were a certain action undertaken, and by minimizing the ensuing cascade of prediction errors, actually brings the action about. This is achieved, ultimately, by a process involving stretch receptors in the muscles, and orchestrated via the dorsal horn of the spinal cord. The point to notice here is that the process that produces the action is (if this story is correct) just another instance of sensory prediction error minimization. The error signal that drives the flow of motor action is simply the difference between current and predicted proprioceptive signals. These signals track states of the body, just as other sensory signals track states of the world. But the body itself is 'known' only via those (and other) sensory signals. So here too, Howhy argues, we encounter the evidentiary boundary. A symptom of this is that as long as the flow of sensory information remains the same, the perceived world (the world insofar as it can be known to the agent) must remain the same.

Our agentive access to the world is thus bounded by the prediction error minimizing routine as it is applied to the flow of interoceptive, exteroceptive, and proprioceptive signals. The upshot, according to Hohwy, is a firm and neurocentric boundary. The PP model, Hohwy concludes:

> "..tells us how neurocentric we should be: the mind begins where sensory input is delivered through exteroceptive, proprioceptive and interoceptive receptors and ends where proprioceptive predictions are delivered, mainly in the spinal cord." (Hohwy, 2014) p. 18

It is for this reason that PP/PEM, as Hohwy constructs it, is claimed to be inimical to many of the core claims associated with recent work on the embodied mind. Thus we read that:

> "PEM should make us resist conceptions of [the mind-world] relation on which the mind is in some fundamental way porous to the world, or viewed as embodied, extended or enactive. Instead, the mind appears to be secluded from the world, it seems to be more neurocentrically skull-bound than embodied or extended, and action itself is more an inferential process on sensory input than enactive coupling with the body and environment" (Hohwy, 2014) p. 1

In the next several sections (sections 3-6) I cast doubt upon this neurocentric reconstruction before proceeding (sections 7-9) to describe and motivate an alternative vision.

3. The Ambiguous Appeal to Inference

Howhy (Hohwy, 2013) pp. 219-221, (Hohwy, 2014)) offers a variety of interlocking considerations meant to support the vision of a secluded, neurocentric mind. The first, and simplest, is the observation (section 2 above) that prediction error minimizing routines are defined over sensory signals so that "from inside the skull the brain has to infer the hidden causes of its sensory input" ((Hohwy, 2013) p.220). The second ties that observation to a traditional form of skeptical threat, and argues that such openness to skepticism is diagnostic of a more traditional ('disembodied') account of mind. The third highlights the presence and importance of an 'explanatory-evidential circle' and argues that this forces us to define a firm evidentiary boundary behind which 'the mind' operates.

As will become apparent, these three arguments are all variants on a single theme, and as such the boundaries between them are not always clear. That theme is one of inferential seclusion – the mind, it is argued, is that which operates behind the veil of transduced sensory information, inferring complex hidden causes as the best explanation of changing (and partially self-induced) patterns of sensory stimulation. There is (as we shall see) a certain sense in which this is correct. But it is important not to over-intellectualize either this process, or the hidden causes themselves. In a wide range of cases, the process that Hohwy describes as one of secluded inference delivers nothing other than an implementation of the kinds of closely coupled perception-action routine highlighted by work on the embodied mind. And these closely-coupled routines provide a kind of step-ladder, I shall later argue, to various kinds of cognitive extension – a step-ladder to what (Clark & Chalmers, 1998) dubbed 'the extended mind'.

Thus consider the first, and most basic, of the arguments mentioned above. This argument points to the role of the brain as a probabilistically-inflected inference engine, constantly attempting to meet the incoming sensory barrage with a stream of matching downwards predictions. According to Hohwy, the operation of such a process induces a disconnect between the mind and the world as it is modeled by the mind. The result is said to be:

> "a schism between the prediction-generating models of the brain and the modeled states of affairs in the world"

By contrast, Howhy adds:

> "Views of mind and cognition that emphasize openness, embodiment, and active extension into the environment seem to be biased against this inferential conception of the mind" Both quotes from (Hohwy, 2014) p.5.

Such views are not, of course, biased against the (surely unassailable) claim that *something* important is being done by the brain when agents engage their worlds in the kinds of ways distinctive of flexible, adaptive, intelligent response. So where might the putative tension lie? It lies in the notion, repeatedly stressed by Hohwy, that what the brain does is best construed as a form of inference. But here we need to be very careful indeed. For the notion of inference in play here is actually far less demanding than it initially appears.

To see this, consider what was at issue in early debates concerning vision and the embodied mind. Here:

> "The key insight… is that the task of vision is not to build rich inner models of a surrounding 3-D reality, but rather to use visual information efficiently and cheaply in the service of real-world, real-time action. Researchers in animate and interactive vision thus reject what Churchland et al [1994] dub the paradigm of 'pure vision' – the idea…that vision is largely a means of creating a world model rich enough to let us 'throw the world away', allowing reason and thought to be focused upon the inner model instead." (Clark, 1999) p.345

The alternative – pursued by successful research programs in 'active perception'– is to use sensing as a channel allowing us to lock-on to simple invariants in the sensory flow[8]. Used in this way, sensing delivers an action-based *grip* upon the world, rather than a neutral reconstruction apt for detached reasoning. Such a grip may intrinsically involve organismic action, as when (to rehearse a familiar case) the baseball outfielder runs so as to keep the image of the ball stationary on the retina. By thus acting in ways that continuously cancel out any apparent optical acceleration, she ensures (Fink, Foo, & Warren, 2009) that she will be in a position to catch the ball when it descends towards the pitch. In such cases, behavioral success is not the outcome of reasoning defined over a kind of inner replica of the external world. Rather, it is the outcome of perception/action cycles that operate by keeping sensory stimulations within certain bounds. This is the same kind of strategy celebrated by work in ecological psychology showing, for example, how some diving seabirds (gannets) predict time-to-impact according to the relative rate of expansion of the image in the optic array - see (Lee & Reddish, 1981) and discussion in (Tresilian, 1999). What matters for present purposes is that these kinds of strategy are *non-reconstructive*. They do not use sensing, moment-by-moment, to build an inner model that recapitulates the structure and richness of the real-world, and that is thus able to stand-in for that world for the purposes of planning, reasoning, and the guidance of action. Instead, here-and-now behavior is enabled by using sensing in the special way described above – as a channel to enable the organism to *co-ordinate* its behaviors with select aspects of the distal environment.

Such non-reconstructive roles for perception are typically cast in bald opposition to the inferential, secluded vision. Thus, in an important recent treatment, Michael Anderson describes non-reconstructive approaches as an alternative to mainstream (inferential and reconstructive) approaches in which perception is cast as analogous to scientific inference and in which:

> "from incomplete and fragmentary data, one generates hypotheses (or models) for the true nature of the world, which are then tested against and modified in light of further incoming sensory stimulation." (M L Anderson, 2014) p.164

These traditional approaches, Anderson continues, depict cognition as "post-perceptual…...representation-rich, and deeply decoupled from the environment". Importantly, Anderson suggest that this follows because, on the traditional accounts he has in mind:

> "reconstructed representations are what the system fundamentally has to work with; the world, once sieved through our senses, simply provides insufficient information about itself. Our understanding of the nature of the epistemic problem that must be solved, then, drives us to hypothesize a particular kind of….solution" (M L Anderson, 2014) p. 164

Non-reconstructive accounts of the role of sensing, Anderson argues, suggest a viable alternative and one that significantly alters our understanding of our own epistemic situation. Instead of engaging the world on the basis of a rich inner model constructed behind the closed doors of sensing, these non-reconstructive solutions show how to achieve behavioral goals by maintaining a delicate dance between sensing and action.

One signature of this kind of grip-based non-reconstructive dance is that it suggests a reversal of our ordinary way of thinking about the relations between perception and action. Instead of seeing perception as the control of action, it becomes fruitful to think of action as the control of perception (Powers (2005)). Thus (re)-conceived, the problem is:

> "not…choosing the right response in light of a given stimulus but…choosing the right stimulus in light of a given goal" (Anderson, 2014) p.182-3).

In other words, it becomes fruitful to see the outfielder (or diving gannet) as running (or diving) in ways that maintain a signature kind of sensory state or flow.

But, as Hohwy himself correctly notes, there is absolutely nothing in the PP/PEM vision that conflicts either with this vision of actions whose role is to harvest perceptions, or (more generally) with the idea of non-reconstructive strategies as one means of promoting behavioral success. Such strategies are, in fact, very naturally accommodated since the best ways to minimize long-term prediction error will often be action-involving, and since there is an in-built premium upon simpler, more efficient solutions. Thus we read that:

> "It is a mistake to think that just because the brain only does inference, it must build up its internal model like it was a following a sober physics textbook. As long as prediction error is minimized on average and over the long run, it doesn't matter which model is doing it. For this reason a model that predicts linear optical trajectories is entirely feasible and can easily be preferable to a more cumbersome series of computations. This is particularly so if it is a less complex model, with fewer parameters, since prediction error in the long run is helped by minimal complexity." (Hohwy, 2014) p.20

This is revealing. Hohwy here (and elsewhere[9]) concedes that often, the PP/PEM framework will indeed stand opposed to more 'intellectualist' frameworks that depict moment-by-moment behavioral success as the product of inferences defined over rich internal models whose role is to allow us to 'throw away the world'. Instead, the role of the inner model is, in very many daily settings, to spot help spot the contexts in which some more frugal, action-dependent, procedure will work (we return to this hybrid picture in sections 7-9 below – see also Clark (2016)). This means that 'inference', as it functions in the PP/PEM story, is not necessarily defined over internal states that bear richly reconstructive, or symbolic, or propositional contents. It is not defined, in other words, over the contents of an inner realm compelled to stand in for the full richness of the external world. Instead, inference may deliver strategies whose unfolding and success depend delicately and continuously upon the structure and ongoing contributions of the external realm, as exploited by action, intervention, and the varying distribution of attention.

Hohwy frequently speaks of neuronal systems as seeking out the *hypotheses* that best explain the sensory information. But it would be more accurate to describe prediction error minimization as a process that finds the multilevel set of neuronal states that best *accommodate* (as I will now put it) the current sensory barrage. This is preferable to talk of 'finding the right hypothesis' as such talk brings unwanted and potentially misleading 'reconstructive baggage'. Accommodating the current sensory barrage may take many forms, some of which involve low-cost methods of selecting actions that re-shape the sensory signal or maintain it within pre-set bounds. Accommodating the incoming signal thus need not (though it sometimes may) imply settling upon an action-neutral description of the external situation, nor need it imply finding a proposition or set of propositions that best describes or predicts that incoming signals. The task of PP systems is not to infer the best description of the world given the sensory evidence. The fundamental task, using prediction errors as the lever, is to find the neuronal activity patterns that most successfully accommodate (in action, and in readiness for action) current sensory states.

4. Evil Demons (Red Herrings)

Why does Hohwy, despite stressing the importance of a 'non-intellectualist' reading of PP/PEM, insist that it promotes a neurocentric, secluded vision of the mind? The reason seems to be that Hohwy links the secluded, inferential vision to something quite different and (I shall argue) rather alien to much of the discussion in hands-on embodied cognitive science. He links it to the mere possibility of global skepticsm. It is this mere possibility that, in Hohwy's treatment, suffices to establish a robust 'veil of tranduction' which positions the world on the far side of an important, agent-impermeable, evidentiary boundary.

Thus, in response to the suggestion that PP/PEM is consistent with (and indeed actively predicts) the use of fast and frugal strategies that use sensing in the special way described above, Hohwy writes that:

> "..the incoming visual signal drives action but…this driving in fact does rely on a veil of transduction, namely the evidentiary boundary within which there is ample inference, and beyond which lies nothing but inferred causes." (Hohwy, 2014) p.21

To demonstrate this, Hohwy repeatedly invokes the spectre of Cartesian skepticism. But this, it seems to me, is a mere distraction (a red herring). The skeptical claim is simply the claim that, were the play of sensory stimulations being received and (apparently) harvested by the brain to remain fixed, so too would our experience of the world. For all we know, then, our physical bodies might be hanging immobile in some Matrix-like energy web, kept alive and fed whatever sensory stimulations are required to make it seem as if we are running to catch fly-balls, and arguing about the powers of evil demons. But this mere possibility (even if it is accepted) in no way casts doubt upon the key claims associated with work in embodied cognitive science. Consider running to catch the fly-ball. This (in the Matrix/vat) would involve feeding the brain the complex, action-sensitive unfolding sensory streams that would normally ensue were an embodied agent actually running so as to cancel the optical acceleration of the ball. The mere fact that this is what would be required attests, it seems to me, to the veracity of the non-reconstructive account of fly-ball interception!

There remains a genuine sense in which the *experienced world* may be said to be constructed by the brain from behind an evidentiary boundary imposed (currently at least) by the biological senses. As Hohwy elsewhere puts it:

> "What are behind the barrier of sensory input are *hidden* causes that must be inferred. An appeal to action, on the prediction error scheme, reduces to an appeal to inferences about different kinds of patterns of sensory input. If a mad scientist was a hidden common cause of all that sensory input we would have no way of knowing unless she made an independent causal contribution to sensory input." Hohwy (2013) p.220, emphasis in original.

The upshot, we are told, is that:

> "our grasp of the world – the way we mirror its causal structure – is at the mercy of the inferential tools we have internally in the brain" Op Cit, p. 221

But the very most that such skeptical challenges could establish would be a very different sense of 'inferential seclusion' from the one at issue in the debates between reconstructive and non-reconstructive approaches to perception and action. For those debates (the ones about the shape of the perception-action nexus) were not about whether we just might be fooled, by some clever manipulation, into misconstruing our own worldly situation. Instead, they were about how best to understand, from *within* our current scientific perspective, the role of the sensory stream in enabling apt forms of world-engaging action. At issue, as we saw, was the question whether apt actions are always and everywhere computed by using sensing to get enough information into the system to allow it to plot its response by exploring an internally represented recapitulation of the distal world. Non-reconstructive solutions, as the name implies, demonstrate the viability of alternative, computationally frugal, but more behaviorally interactive. They do not imply – nor do they seek to imply – the falsity of the skeptical hypothesis. That, I suggest, is an orthogonal question that demands a full philosophical treatment in its own right[10]. Instead, non-reconstructive (broadly speaking 'ecological') accounts are a promising move in what is first and foremost a very different game - the game of understanding, in the light of our best empirical science, the actual role of perception in a wide variety of fluent behaviors[11].

The image of the mind as secluded behind an inferential curtain is thus itself importantly ambiguous. If it means only that the world, insofar as we know and experience it, is that which is both specified and engaged by the ongoing flow of (partially self-induced) sensory stimulations, then PP/PEM indeed mandates a certain kind of seclusion. Though even there, the question of where to place the boundaries of sensing themselves – and the question whether well-fitted tools and technologies might temporarily or permanently alter the best place to place those bounds – arises (see section 4 below). But seclusion, in this rather limited sense, does not imply the richly reconstructive model of perception according to which our actions are selected by processes of reasoning defined over the contents of rich inner models whose role is to *replace* the external world with a kind of inner simulacrum[12].

The mere fact that neural processing is organized around prediction error minimization routines thus puts no real pressure upon the claim that lies at the heart of recent work on the embodied mind. For what that work most fundamentally rejects is the *richly reconstructive model of perception*. The appearance of conflict arises from ambiguities in the notions of inference and seclusion themselves. For these notions may seem to imply the presence of a rich inner recapitulation of the distal environment, with a consequent *downgrading* of the role of action and upgrading of the role of reasoning defined over that inner model. Nothing in PP/PEM, however, mandates this. On the contrary, PP/PEM strongly suggests that brains like ours will, wherever possible, exploit simple strategies that rely heavily on world-engaging action, delivering new sensory stimulations just-in-time to support behavioral success.

5. Making Space for The Extended Mind.

Hohwy's argument for a secluded, inferential model of mind trades heavily, it seems to me, on the ambiguity between reconstructive and non-reconstructive uses of sensing identified in the previous section. Thus consider the following passage, which aims to put direct pressure on the idea that human minds might be ‘extended minds’ in the sense of (Clark & Chalmers, 1998). The suggestion there was that the true circuits of human cognition might (when certain further conditions are met) include operations and storage realized by bio-external resources such as a smartphone or even a notebook. I shall not attempt to convince the reader of this view here[13]. What I do hope to show, however, is that nothing in the PP/PEM framework should negatively impact the arguments already supposed to favour the conclusion that the machinery of mind can, sometimes, extend out into the extra-neural world.

Commenting on this issue, Hohwy writes that:

> “An agent can grasp and use her phone only because she has a more or less precise and accurate internal representation of the phone, the things in her drawer that may occlude it, and the causal interactions between her fingers, eyes, voice and the states of the phone…..In so far as we can interact with things in the environment, including our bodies and other people and their mental states, we must be modeling them, forming hypotheses about them and their interactions, predicting the next sensory input, assessing the prediction error generated and updating the hypotheses accordingly. In other words, there is reason to think that these states are all hidden causes, situated beyond the evidentiary boundary.” Hohwy (2014) p.11

We can now see what is wrong (or at least subtly misleading) with this diagnosis. It is correct insofar as (assuming the truth of PP/PEM) the interactions between the biological organism and the smartphone are indeed orchestrated by bio-internal processes of prediction error minimization. But it is misleading to suggest that, moment-by-moment, our fast and fluent uses of the smartphone (or any other bio-external resource) require us to command a ‘precise and accurate internal representation of the phone’. Instead – just as in the case of running to catch the fly-ball – what may often be doing the work is a kind of perceptually-maintained motor-informational grip on the world: a low-cost perception-action routine that retrieves the right information just-in-time for use, and that is not in the business of building up a rich inner simulacrum[14].

To see this, reflect that the operations and information stores made available by some bio-external resource may become densely woven into a set of learnt habits: compiled, easily cued motor routines by means of which the intelligent agent deals with her (wider) world. Within such a weave, there is no clock governing the brain’s exchanges with the world (it is not the case that all the inner stuff is done at t1, and calls to the world happen at t2). Nor is it the case that all the interactions between the biological and bio-external resource are launched and routed through the slow, serial bottleneck in which conscious attention and/or the agent's intentions are used to guide deliberate action. Instead, as in skilled performance more generally, activity in the brain becomes dovetailed with multiple sub-personally orchestrated 'calls to the world' accomplished by embodied action. Instead, such cases involve a temporary coalition of unfolding internal processes, each of which may directly issue, at differing time-scales, calls both to other inner processes and to outward-looping 'epistemic acts' that harvest new information, or call upon new information-transforming operations, just-in-time to keep the process rolling (see (Kirsh & Maglio, 1994)). The brain is not required explicitly to *represent* the availability of such and such information/operations from any given internal or external location. Instead, it simply deploys a problem-solving routine whose fine structure has been selected (by learning and practice) so as to assume the easy availability of such and such information or the easy accomplishment of such and such a useful data-transformation, from (for example) such and such a visual location via the performance of such-and-such a gross motor action. Similarly, when our brains detect a sudden flash and our eyes automatically saccade in that direction, the motor routine embodies a kind of unrepresented commitment to the effect that we may gain useful (perhaps life-saving) information by such a rapid saccade.

The effect of extended problem-solving practice is thus to install a kind of motor-informational weave such that repeated calls to bio-external resources become built-in to the very heart of many of our daily cognitive routines. But to repeat: such calls need not depend on (consciously or unconsciously) representing the fact that such-and-such information is available by such-and-such a motor act. Applied to the smartphone case, the moral is that to use the phone in fluent, semi-automatic ways our brains don't need to model or represent it in any equivalently rich way. So the idea that *everything that matters* about the phone when it is folded into a fluent problem-solving flow is exhausted by the way it turns up as an 'inferred cause' in my brain's PGM (thus placing it behind the evidentiary boundary imposed by our sensory systems) is simply false.

The idea that the phone – in it's potential role as part of the machinery of human cognition - is merely an 'inferred cause' hidden behind the evidentiary boundary imposed by our sensory systems thus loses its sting. This is not to deny that insofar as we *experience and understand* the phone, that experience and understanding constructs the phone as an inferred cause (on a par with tables, chairs, and other minds). But that doesn't in itself imply that the phone is not – also, simultaneously – acting as part of our own 'extended cognitive architecture'. The same is true, after all, of my own brain! Insofar as I know about and (perhaps via sophisticated imaging techniques) perceptually experience my own brain, that's because it too has turned up in my model of the world, as another 'inferred cause' constructed from behind the evidentiary veil imposed by the sensory surfaces. But this does not preclude its also functioning as *part* of my cognitive equipment – a part, moreover, that can contribute to my cognitive processing in myriad ways that are not exhausted (thankfully) by what I know about it. A corollary of this is that the smartphone or other bio-external equipment could, in principle, participate in an agent's cognitive processing without that participation being exhausted by, or limited by, or moment-by-moment constituted by, what that agent - or even that agent's entire internal probabilistic generative model - knows or encodes about it. In such cases the bio-external resource participates in a dense, sub-personal, motor-informational weave that implements cognitive skills without replicating them 'in the head'.

This highlights something crucial that is often overlooked in discussions of the extended mind. It is that sensing, given such a picture, plays two very different roles with respect to the smartphone, or any other bio-external resource. Sometimes, sensing plays the standard role of enabling an agent to see and think about the resource – as when we ask ourselves which model smartphone to purchase, and go to the store to examine the leading candidates. At other times (once the phone has been assimilated into patterns of fluent, unreflective use) it plays a role more like that of an inner information flow within the brain – so that an act of sensing is now more like an information transferring relay within a larger information processing whole. Plugged into the larger fabric of existing arguments in favour of extended cognition all this suggests that smartphones and other well-fitted 'cognitive prosthetics' can indeed participate in episodes of cognitive processing. What seems to matter in these cases is that the acts and operations that they make available become intimately sub-personally dovetailed with the acts and operations provided by neuronal and gross-bodily resources, so that it is the whole transient ensemble that is called upon, fluently and automatically, in the service of problem-solving success[15].

At the very least, we should allow that neural and extra-neural operations can become inextricably interwoven within the kinds of skilled intelligent commerce most characteristic of human cognitive success. The existence of such dense, sub-personal, supra-representational motor-informational weaves should already (independently of arguments for or against extended cognition) give us sufficient cause to reject the suggestion that PP leads to a secluded, disembodied, neurocentric, vision of mind. Or does it?

6. Self-Evidencing Systems

The neurocentric, skull-bound vision of mind also follows, or so (Hohwy, 2014) argues, from a certain implication of the PP/PEM story. It follows, we are told, from the implied vision of the brain as a *self-evidencing system*. Self-evidencing (Hempel, 1965) occurs when a hypothesis best explains some piece of evidence and, in virtue of that explanatory success, thereby provides evidence for its own truth or correctness. In such cases, the occurrence of the evidence is best explained by the hypothesis but the fact that the evidence occurs at all is used to lend support to the hypothesis itself. To use a common example, my lack of study may be offered as an explanation of why I failed the exam, while my failing the exam might reasonably be offered as evidence for my lack of study. This can sound unacceptably circular. Despite this, we make use of such forms of reasoning daily, and in ways that can be quite epistemically innocent. Thus Peter Lipton notes that:

> "Self-evidencing explanations are common, in part because we often infer that a hypothesis is correct precisely because it would, if correct, provide a good explanation of the evidence. Seeing the disemboweled teddy bear on the floor, with its stuffing strewn throughout the living room, I infer that Rex has misbehaved again. Rex's actions provide an excellent if discouraging explanation of the scene before me, and this is so even though that scene is my only direct evidence that the misbehaviour took place. To take a more scientific and less destructive example, the velocity of recession of a galaxy explains the redshift of its characteristic spectrum, even if the observation of that shift is an essential part of the scientist's evidence that the galaxy is indeed receding at that the specified velocity." P. Lipton in (Hon & Rakover, 2013) p. 44-5

The scientific hypothesis concerning the velocity of recession of the galaxy is here self-evidencing. The hypothesis (that the velocity of recession is such-and –such) explains the redshift, and the observation of the redshift provides evidence of that very velocity of recession.

Something similar may be claimed for PP/PEM. Thus Friston writes:

> "I model myself as embodied in my environment and harvest sensory evidence for that model. If I am what I model, then confirmatory evidence will be available. If I am not, then I will experience things that are incompatible with my (hypothetical) existence. And, after a short period, will cease to exist in my present form" (K. Friston, 2011) p. 117

Friston makes clear, however (K. Friston, 2011, 2013a, 2013b) that this talk of 'the model' is meant to pick out the whole embodied agent. The whole embodied agent, then, is the full 'model' whose evidence is maximized by its own success (persistence). The generative model that is implemented by wiring and activity patterns in the brain is thus treated as contributing to, but by no means exhausting, the overall 'embodied model' that "distils and embodies causal structure in its local environment" p. 89). In this (slightly unusual) sense:

> "… an agent does not have a model of its world – it is a model. In other words, the form, structure, and states of our embodied brains do not contain a model of the sensorium– they are that model." (K. Friston, 2013a) p. 32

This, I submit, is the primary locus for 'self-evidencing' in the PP/PEM story that:

> "takes the existence of agents as its starting point and concludes that each phenotype or agent embodies an optimal model of its econiche" (K. Friston, 2011) p.89

Hohwy, in the treatment of self-evidencing, fails to foreground this important wrinkle. In Hohwy's treatment, it is the notion of the self-evidencing *brain*, rather than any notion of a *self-evidencing embodied agent*, that is supposed to impose the evidentiary curtain that ushers internalism back onto the cognitive arena. Once look at the larger (fully embodied) story, things start to look rather different.

To see this, notice that the PP/PEM accounts are often presented as manifestations of an even more general principle known as 'free energy minimization'[16]. But the notions of inference and model, as they are used within this larger information-theoretic framework, are *extremely* weak. For example, (K. Friston, Levin, Sengupta, & Pezzulo, 2015) using the free energy framework to describe the way cells migrate and differentiate during embryogenesis, comment that:

> "If each cell… minimizes variational free energy then it should, in principle, come to *infer* its unique place in the ensemble and behave accordingly. This is guaranteed because the minimum of variational free energy is obtained when each cell is in a unique location and has correctly inferred its place. At this point, it will express the appropriate signals and fulfil the predictions of all other cells; thereby, maximizing the evidence for its *model* of the ensemble (and minimizing the free energy of the ensemble" (K. Friston, Levin, et al., 2015) p.2, my emphases

Whatever the use of the terms 'infer' and 'model' mean in these low-level free energy minimization accounts, they does not seem to imply the presence of inner models or content-bearing states of the kinds imagined in traditional cognitive science. Instead, what are picked out seem to be physical processes defined over states that do not bear contents at all – neither richly reconstructive *nor* of any more 'action-oriented' kind[17].

How should we understand the notion of 'self-evidencing' in these kinds of case? Consider a very simple creature, such as a bacterium. Friston will say of such a creature that its very existence provides evidence for *itself* considered as an 'embodied model' of the organism-salient environment. This is because even in the case of the bacterium, its inner states "must entail a generative model of its world whose free energy is minimized by perception and action". This is true, we are told "whether you are an E. coli or an evangelist. Because free-energy is a function of sensations and internal states it is, in essence, an attribute of an embodied inference." (both quotes from (K. Friston, 2011) p.117).

In the E. Coli case there is thus self-evidencing aplenty. But this in no way detracts from the obvious fact that much that matters about the E. Coli is clearly 'extra-neural' and may crucially involve (for example) the placement, length, and flexibility of the motion-enabling flagellum. If we (as external theorists) wish to explain and understand the way in which E. Coli remains within its specialized window of viability, temporarily resisting the second law of thermodynamics, this suggest that we really do need to treat the *whole embodied bacterium* as the free–energy minimizing 'model'. In that extended sense, bodily forms, sensors, sensor placements, and structures (just like gross neuroanatomical forms and structures – see (K. Friston, 2011) must be considered as parts of the 'embodied model' itself.

Let us accept that, as far as the brain itself is concerned, all the evidence it (the brain) ever gets is evidence mediated by perturbations of our sensory surfaces. We can now see that it does not follow that every change in the total free-energy minimizing 'embodied model' that (in Friston's extended sense) *we are* must be mediated by, or result in, any change in an *inner model* supported by the brain/CNS. The broader free energy story to which Hohwy appeals thus fails to support any strong conclusions concerning neuro-inferential seclusion. For the 'model' in question is not (not exclusively, and sometimes, as in the case of E. Coli, not at all) an inner content-bearing model bounded by activity at the sensory surfaces. Instead, the model is the whole embodied organism whose gross bodily forms and features are themselves free-energy minimizing devices.

Interestingly, Hohwy himself repeatedly notices – and even stresses – the importance of the whole organism. But he does not mark it as complicating (perhaps even undermining) his own arguments for neurocentric seclusion. The reason, I suspect, turns once again upon perceived implications of the mere possibility of global skepticism. Thus Hohwy may reply that, from the creature's own perspective, the world *as it knows it* must be fully specified by explicit processes of prediction error minimization applied to activity patterns at its sensory surfaces (as they occur against an evolving backdrop of priors and precision estimations). It is this, we are told, that requires us to embrace the bare possibility of global skepticism.

I do not wish to enter into these hotly contested skeptical debates here (for my own take on such possibilities, see Clark (2005) and the discussion in Chalmers (2005), both of which appear in (Grau, 2005))[18]. But what should at least be clear is that the full free energy story depicts an agent-world relation that is far richer than any relation between explicit neurally realized inner models and the larger environment. The full agent–world relation foregrounded by the free energy accounts is *in no way hermetically bounded by the agent's own explicit neural constructs*. The mere possibility that we may be deeply and permanently misled about our apparent surroundings is thus orthogonal to the real question before us, which concerns the most likely shape of the *actual* mind-world relation, assuming the existence of the creature and the truth of the PP/free energy framework.

To sum up, the argument from the self-evidencing brain trades, I suggest, upon another ambiguity - one that is very closely related to that explored in the previous section. This time, it is the ambiguity between the common notion of inner models and a much broader sense of 'model', sometimes flagged by talk of 'the embodied model'. In the more restricted sense, what matters (for understanding mind and intelligent, adaptive response) is only the ways sensory inputs impact an articulated inner model. But in the broader sense, whole embodied agents are the models and 'free energy' is minimized by all manner of adaptive tricks and ploys. These may (but need not) include the use of complex explicit inner models. But they also include the use of very simple inner models, facts about sensor-types, gross morphology, sensor placement, and the very materials of which the organism is built. This means that it is not brains (or brain-based inner models) but whole embodied agents that are, in the relevant sense, self-evidencing[19].

7. Efficiency: A Dilemma for the Neurocentric Vision

So far, our discussion has been mostly negative. Nothing in the PEM story, so the argument goes, should force us to accept a secluded, neurocentric vision in which inner constructs do all the heavy lifting. It is now time to think more positively and explore an alternative account. That account depicts predictive processing as a thoroughly dynamical story that highlights self-organization and complex brain-body-world interactions, and that thus provides the perfect partner for work on embodied, extended and enactive cognition.

The key concept here is efficiency. Efficiency (see e.g. Barlow (1959), (Olshausen & Field, 1996)) is intuitively the opposite of redundancy and excess. A representational scheme is efficient if it uses only the minimal resources necessary to capture the regularities that matter for driving behavior. In general, this means finding a model that, when confronted with new sensory data, need only update a few parameters to account for (to 'explain away') the gross sensory signal. A model that is rich enough to capture the regularities important for selecting behavior, but that requires alterations to very few of its critical states to explain the sensory signal, has low complexity in this sense. Such models provide what ((Hobson & Friston, 2014) p.23) describe as "accurate but parsimonious explanations for sensations". The goal of the predictive brain, in other words, is to command models that track organism-salient patterns without relying upon more parameters than are positively required to do the job. Systems that fit data accurately (using the minimum number of parameters) are efficient modelers of their world. A system that uses a large number of parameters to explain the same data is not a 'more accurate' modeler of its world. On the contrary, the result will often be 'over-fitting' of the model to the observed data, some of which turns out to be merely 'noise' or random fluctuations rather than informative signal. This is nicely dramatized in Feldman's (2013) p.15) discussion of the 'Lord's Prior" where this rather mischievously names the misleading idea (roundly rejected by Feldman) that "the optimal Bayesian observer is correctly tuned when its priors match those objectively in force in the environment".

The deepest problems with such a notion emerge as soon as we reflect that active agents are not, at root, simply trying to model the data so much as to come up with recipes for choosing which data to sample next - recipes for acting appropriately in the world. The Optical Acceleration Cancellation procedure described in section 3 provides a nice example, since it combines low complexity (few parameters) with high behavioral leverage. Commenting on the OAC model, Anderson writes that:

> "It is important to underline how big a blow this apparently simple finding is for the traditional view. In order to solve this fairly complex perception-action coordination problem, people do not appear to be reconstructing or otherwise representing the flight path of the ball, nor generating predictive hypotheses. Instead they are acting continuously in real time so as to achieve a particular perception." (M L Anderson, 2014) p.185

Anderson clearly takes the OAC strategy to be representative of a large and rich space of alternatives to traditional views that involve complex 'epistemic mediators'. And it is easy to see what he is getting at. For a system using the OAC strategy is not engaging in a process of inner reconstruction whose goal is to enable the organism to solve the problem 'from the inside', merely by operating upon a rich suite of inner encodings. Instead, this is a problem solution that requires (all those marginal skeptical possibilities notwithstanding) behavior to unfold courtesy of close couplings between body and world – couplings that the brain maintains by the simple device of ensuring that the flow of sensation remains within certain bounds. Action is here the control of perception, and perception is not about building up a rich inner model but about maintaining the rolling perception-action cycle that solves the problem.

Anderson claims that the brain, within this unfolding cycle, is not 'generating predictive hypotheses'. But Hohwy, as we saw earlier, expressly endorses the view that even in cases such as these, the PP/PEM story applies. Hohwy is right. The relevant distinction is *not* between the presence or absence of a strategy rooted in neural prediction (and the resolution of neural prediction error). For the OAC strategy is easily implemented using prediction error minimizing techniques – simply treat as salient (highly weighted) all and only the prediction errors associated with optical acceleration of the ball. Cancelling those errors by action is, in fact, an excellent example of an action-based

PP/PEM strategy – the kind of strategy that should be favored by the considerations of efficiency and minimal modeling just noted[20].

Anderson is right, however, to stress the epistemic radicalism of the OAC story. For OAC-style solutions are, just as Anderson insists, deeply non-reconstructive. They are 'grip-based' stories that undermine the traditional vision according to which cognition is accomplished by operations defined over a kind of picture-perfect inner recapitulation of the external world. By stressing coupled unfoldings and action as the control of perception, such stories reveal cognition as a world-engaged, action-oriented process. Such accounts stand opposed to visions of epistemic isolation.

The upshot is a kind of dilemma for Howhy's defense of the neurocentric vision. For Hohwy, as we have seen, wants to *combine* a staunchly 'non-intellectualist' reading of PP/PEM with much more conservative claims concerning epistemic insulation. But this is not a consistent combination. Hohwy's vision of epistemic isolation would follow only if predictive brains calculated behavioral responses using richly reconstructive inner models. To be sure, some critics of the predictive vision (such as (Michael L Anderson & Chemero, 2013) mistakenly *identify* genuinely prediction-based accounts with either symbolic-inferential or richly reconstructive approaches[21]. Hohwy, by contrast, clearly believes that PP/PEM accounts easily encompass the kinds of efficient and elegant solution familiar from work in embodied cognitive science. This is correct. But that means it is not, by any stretch of the imagination, an isolationist brain-bound treatment. Instead it is one that places coupled unfoldings, the active embodied agent, and the enabling environment centre-stage. The vision of neurocentric isolation is thus incompatible with Hohwy's careful and important recognition of the wide range of non-reconstructive solutions implied by work on the predictive brain.

8. Transient Extended Cognitive Systems

The PEM/PP framework, with its deep commitment to efficiency in neural processing, has the potential to illuminate large swathes of work in embodied cognitive science. But to fully appreciate that potential we must first notice another key ingredient in the predictive processing economy. That ingredient is the capacity to deliver moment-by-moment reconfigurations of patterns of 'effective connectivity' within the brain.

To see how this works, consider first that (within the PP/PEM framework) first-order probabilistic expectations are intertwined with context-varying assessments of the reliability and salience of different bodies of information. These second-order estimations of reliability and salience determine the weighting (or 'precision') given to different aspects of the prediction error signal at different levels of processing. The primary effect of this is to systematically vary the relative influence of top-down versus bottom-up information by increasing the gain (intuitively, increasing the 'volume') on selected error units. This enables the well-calibrated perceiver to rely more on the sensory evidence when conditions are favorable, and to allow top-down expectations to play a larger role when the sensory signal is noisy or compromised. It also allows for greater reliance on select modalities (or more specific aspects of the sensory signal) according to variations in context and task demand. Precision estimations thus provide a powerful and delicate tool for putting stored knowledge to use in ways that vary with task and context. What this really amounts to is something quite spectacular. For variable precision-weightings are thus sculpting the patterns of 'effective connectivity' that vary internal (and thus, as we'll next see, external) flows of influence and information according to task and context[22].

With these tools in hand, let's revisit the outfielder's problem described earlier. In such a cases, active neural predictions and simple, rapidly-processed perceptual cues work together to select a pattern of precision-weightings for different prediction error signals. This creates a transient web of effective connectivity (a simple, temporary, distributed circuit or what Anderson (2014 p. 94) dubs a TALoN – Transiently Assembled Local Neural Subsystem) and, within that circuit, sets the balance between top-down and bottom-up modes of influence. The temporary task of visual sensing, in this context, becomes that of cancelling the optical acceleration of the fly ball (hence giving high weighting to prediction errors associated with cancelling the vertical acceleration of the ball's optical projection). In this way, apt precision weightings select a pre-learnt, fast, low-cost strategy for solving the problem. Contextually recruited patterns of precision weighting thus accomplish a form of set-selection or strategy switching – an effect that has been practically demonstrated in some simple simulations of cued reaching (K. J. Friston et al., 2012)).

Such solutions assume that slower processes of learning and adaptive plasticity have already sculpted patterns of neural connectivity in ways that make the low-cost (e.g. Optical Acceleration Cancellation) strategy available. But this is unproblematic. It can be motivated in general terms by the drive towards energetic efficiency, and implemented using processes of prediction error minimization at many time-scales. Such processes range all the way from the slow learning of the child baseball player, to the faster online adaptation of the pro-player factoring in (during a match) the changing specifics of wind conditions and the play of opposing batters. The upshot will be a highly tuned system in which multi-scale predictive learning and rapidly-recruited patterns of precision-weighting conspire to control changing patterns of effective connectivity – in this case, making available a fast, low-cost strategy for solving the problem.

Putting all this together yields a complex but rewarding picture in which bedrock processes of predictive learning slowly install models that include precision expectations allowing patterns of effective connectivity to be selected 'on the fly'. Such patterns in turn allow fast, knowledge-sparse modes of response to be recruited and nuanced according to current context. But more complex (intuitively more inferential and 'model-rich') strategies may also involve simplifications and approximations. A nice example is work by (Battaglia, Hamrick, & Tenenbaum, 2013) on 'intuitive physics'. Human agents are able to make rapid inferences about the physical behavior of ordinary objects. Such inferences might include spotting that the pile of books or washing-up is unstable and at risk of toppling over, or that a lightly brushed object is going to fall and hit some other object. Underlying that capacity, Battaglia et al suggest, may be a probabilistic scene simulator (a probabilistic generative model) able to deliver rapid verdicts on the basis of partial, noisy information. Such a simulator does not rely upon propositional rules but rather upon "quantitative aspects and uncertainties of object's geometry, motions, and force dynamics" (op cit p.18327). Approximate solutions such as these reflect what Gershman & Daw (2011, p.307) describe as a kind of "meta-optimization over the costs (e.g. extra computation) of maintaining [a] full representation relative to its benefits". The deepest explanation for the neural intermingling of perception, action, and utility may, Gershman and Daw (op cit, p.308) suggest, lie right there, in adaptive pressure to find and deploy representational forms and statistical approximations that "concentrate their density in regions of high utility". The upshot is a kind of meta-Bayesian determination of what to represent, and of when, and how, to represent it. Meta-Bayesian agents like this will use the most efficient strategy that is good enough to do the job, and that is currently available within the reconfigurable flow. Dealing with a complex time-pressured world thus demands the use of many strategies, ranging from very simple heuristics to more complex structures of interacting approximations. That diverse landscape may, however, form part of an overarching uncertainty-based cognitive eco-system – an eco-system within which these many strategies emerge, dissolve, and interact. The PP/PEM architecture is thus dynamically self-reconfiguring, constantly engaging actions that yield new inputs that recruit new strategies in a kind of rolling cycle.

Such architectures are ideally positioned to support the kinds of 'motor-informational weave' highlighted in section 4 above. To see this, reflect that known external (e.g. environmental) operations provide – by partly constituting - additional strategies apt for the kind of 'meta-model-based' selection just described. This is because actions that engage and exploit specific external resources will now be selected in just the same manner as the inner coalitions of neural resources themselves. For example, consider the case where salient high-precision information is available by the use of some bio-external device, such as a laptop or smartphone. The core routine that selects actions to reduce prediction error will now select actions that invoke the bio-external resource. Invoking a bio-external resource, and moving our own effectors and sensors to yield high-quality task-relevant information, are here expressions of the very same underlying strategy: one that reflects our brain's best (sub-personal) estimates of where and when reliable, task-relevant information is available.

As a further illustration, consider work by (Pezzulo, Rigoli, & Chersi, 2013). Here, a so-called 'Mixed Instrumental Controller' determines whether to choose an action based upon a set of simple, pre-computed ('cached') values, or by running a mental simulation enabling a more flexible, model-based assessment of the desirability, or otherwise, of actually performing the action. The mixed controller computes the 'value of information' selecting the more informative (but costly) model-based option only when that value is sufficiently high. Mental simulation, in those cases, then produces new reward expectancies that can determine current action by updating the values used to determine choice. We can think of this as a mechanism that, moment-by-moment, determines whether to exploit simple, already-cached routines or to explore a richer set of possibilities using some form of mental simulation. It is easy to imagine a version of the mixed controller that determines (on the basis of past experience) the value of the information that it believes would be made available by some kind of cognitive extension, such as the manipulation of an abacus, a smartphone, or a physical model. Deciding when to rest content with a simple cached strategy, when to deploy a more costly mental simulation, and when to exploit the environment itself as a cognitive resource are thus all options apt for the same kind of 'meta-Bayesian' model-based resolution.

Seen from this perspective, the selection of task-specific *inner* neural coalitions within an interaction-dominated PP economy is entirely on a par with the selection of task-specific *neural-bodily-worldly* ensembles. PP thus delivers a perfect fit with the vision of the brain as a dynamical engine whose key role is to initiate and maintain valuable patterns of embodied interaction ('grip') with a richly structured environment. This, of course, is precisely the radical vision of the embodied mind that Hohwy depicts PP as opposing. It is the vision championed, for example, in Anderson (2014), who summarizes is as implying that:

> "the brain is fundamentally action-oriented, and specializes in managing the organism's interactions with the world; and….the brain achieves its functions by assembling the right functional coalitions between both neural and extra-neural partners, including supporting interaction with external artifacts including symbolic ones—for cognitive ends." Anderson (2014) p. 302

The recruitment and use of extended (brain-body-world) problem-solving ensembles now turns out to obey many of the same basic rules, and reflects many of the same basic normative principles (balancing efficacy and efficiency, and reflecting complex precision estimations) as does the recruitment of temporary *inner* coalitions bound by effective connectivity. In each case, what is selected is a temporary problem-solving ensemble (a transient extended cognitive system (Clark (2008)) or 'task-specific device' (Bingham, 1988) recruited as a function of context-varying estimations of uncertainty[23]. Such temporary ensembles emerge and are deployed within iterated cycles in which perceptuo-motor routines deliver new inputs that recruit new transient ensembles of resources. It is these rolling cycles that most clearly characterize human cognition in the wild. Within such cycles, arbitrarily complex amounts of 'leaning on the world' may become progressively folded in, expanding our practical cognitive capacities by offloading work from brain to (non-neural) body, and from organism to (physical, social, and technological) world. What PP makes unusually clear is that it is these rolling cycles that the neural economy constantly (and not just in the special cases involving mind-extending tools and technologies) serves. As these complex engagements unfold, no inner homunculus oversees the repeated soft-assembly of the distributed problem-solving ensembles that result. Instead, such ensembles emerge and dissolve in ways determined by the progressive reduction of precise, high-quality, prediction error. Organismically salient (high precision) prediction error may thus be the glue that, via its expressions in action, binds elements from brain, body, and world into temporary problem-solving wholes.

9. Conclusions: Prediction-Based Intermingling

Work on embodied, extended, and enactive cognition can seem to point in a rather different direction to work on the predictive brain. The perceived conflict is, however, illusory. Once we consider the role of prediction in the genesis and unfolding of action, the picture alters dramatically. For predictive processing results in the creation and deployment of 'pragmatic action-oriented representations': inner states tailored to the production of good online control rather than aiming for rich reconstructive mirroring of some action-independent world. Instead, neural processing delivers a grip upon a world of possibilities for action and intervention. Perception delivers a world parsed for action, while action harvests the perceptual flows that secure both epistemic and practical success[24].

Predictive processing's full resonance with work in embodied cognition can only be appreciated, however, once an additional ingredient is brought into play. That ingredient is the capacity to use variable 'precision weighting' to sculpt patterns of effective connectivity within the brain, thereby selecting actions that recruit the simplest brain-body-world circuits that can reliably support a target behavior. This neatly accommodates frugal 'sensing-for-coupling'-style solutions of the kind celebrated by work in ecological psychology. But better still, it accommodates those solutions within the systematic and empowering context of a fluid, re-configurable economy in which the use of rich, knowledge-based strategies and the use of fast, frugal procedures are merely different expressions of a common uncertainty-estimating mechanism. Thanks to that mechanism, changing ensembles of inner and outer resources are repeatedly recruited, forming and dissolving in ways determined by external context, current needs, and ongoing (sub-personal) estimations of our own uncertainty. The threat of neurocentric seclusion is thus fully and satisfyingly averted. What remains is a vision of dense, but fluid, intermingling in which brain, body, and environment appear as "mutually embedded systems" (Thompson & Varela, 2001 p.423) harmonized in the service of situated success.

• Thanks to Mark Miller and Chris Burr for helpful comments on these ideas. This work was supported in part by the AHRC funded 'Extended Knowledge' project, based at the Eidyn research centre, University of Edinburgh.

References

Aertsen, A., Bonhöffer, T., & Krüger, J. (1987). Coherent activity in neuronal populations: Analysis and interpretation. In E. R. Caianiello (Ed.), *Physics of cognitive processes* pp. 1-34. Singapore World Scientific Publishing.

Adams, F., & Aizawa, K. (2008). *The Bounds of Cognition.* (John Wiley and Sons)

Anderson, M. L. (2014). *After Phrenology: Neural reuse and the interactive brain*. MIT Press.

Anderson, M. L., & Chemero, T. (2013). The problem with brain GUTs: Conflation of different senses of "prediction" threatens metaphysical disaster. *Behavioral and Brain Sciences*, *36*(03), 204–205.

Anderson, M. L., Richardson, M., & Chemero, A. (2012). Eroding the boundaries of cognition: Implications of embodiment. *Topics in Cognitive Science*, *4*(4), 717–730.

Ballard, D. (1991). Animate vision. *Artificial Intelligence*, *48*, 57–86.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*, 695–711.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18327–32. doi:10.1073/pnas.1306572110

Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, *4*(3), 91–99.

Bingham, G. P. (1988). Task-specific devices and the perceptual bottleneck. *Human Movement Science*. doi:10.1016/0167-9457(88)90013-9

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00599

Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, *4*(March), 25. doi:10.3389/fnhum.2010.00025

Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J., & Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(27), 11262–11267. doi.org/10.1073/pnas.1011284108.

Carey, S. (2009). *The origin of concepts*. NY: Oxford University Press.

Castiello, U. (1999). Mechanisms of selection for the control of hand action. *Trends in Cognitive Sciences*. doi:10.1016/S1364-6613(99)01346-7

Chalmers, D (2005) The Matrix as Metaphysics, in R. Grau (ed) *Philosophers Explore The Matrix,* Oxford University Press, NY

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, MA: MIT Press.

Churchland, P. S., Ramachandran, V. S., & Sejnowski, T. J. (1994). A Critique of Pure Vision, In: C. Koch and J. Davis (Eds.), Large-Scale Neuronal Theories of the Brain, 1–25.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*, 1585–1599. doi:10.1098/rstb.2007.2054

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron*, *45*, 801–814. doi:10.1016/j.neuron.2005.01.027

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, *33*, 269–298. doi:10.1146/annurev.neuro.051508.135409

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. MIT press.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, *3*(9), 345–351.

Clark, A. (2003). *Natural born cyborgs: Minds, technologies, and the future of human intelligence.*

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension* (Oxford University Press, NY)

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. doi:10.1017/S0140525X12000477

Clark, A (2016) *Surfing Uncertainty: Prediction, Action, and the Embodied Mind* (Oxford University Press, NY)

Clark, A., & Chalmers, D. (1998) The Extended Mind *Analysis* 58: 1: pp. 7-19.

Cocchi, L., Harding, I. H., Lord, A., Pantelis, C., Yucel, M., & Zalesky, A. (2014). Disruption of structure–function coupling in the schizophrenia connectome. *NeuroImage : Clinical*, *4*, 779–787. doi:10.1016/j.nicl.2014.05.004

Cole, M. W., Anticevic, A., Repovs, G., & Barch, D. (2011). Variable Global Dysconnectivity and Individual Differences in Schizophrenia. *Biological Psychiatry*, *70*(1), 43–50. doi:10.1016/j.biopsych.2011.02.010

Damasio, A., & Damasio, H. (1994). Cortical systems for retrieval of concrete knowledge: The convergence zone framework. In C. Koch (Ed.), *Large-scale neuronal theories of the brain* (pp. 61–74). Cambridge, MA: MIT Press.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, *39*, 1–38.

Edelman, G. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York: Basic Books.

Edelman, G., & Mountcastle, V. (1978). *The mindful brain: Cortical organization and the group-selective theory of higher brain function*. Cambridge, MA: MIT Press

Elman, J.L. (2005). Connectionist models of cognitive development: Where next? *Trends in Cognitive Science,* 9, 111-117.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews*, 2, 704–716.

Feldman, J. (2013). Tuning your priors to the world. *Topics in Cognitive Science*, *5*(1), 13–34. doi:10.1111/tops.12003

Fink, P. W., Foo, P. S., & Warren, W. H. (2009). Catching fl y balls in virtual reality : A critical test of the out fielder problem *Journal of Vision*, *9*(13):14, 1–8.

Fornito, A., & Harrison, B. J. (2012). Brain Connectivity and Mental Illness. *Frontiers in Psychiatry*, *3*, 72. doi:10.3389/fpsyt.2012.00072

Friston, K. (1995). Functional and effective connectivity in neuroimaging: A synthesis. *Hum. Brain Mapp.*, *2*, 56–78.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 815–36. doi:10.1098/rstb.2005.1622

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

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–38. doi:10.1038/nrn2787

Friston, K. (2011). Embodied inference: or "I think therefore I am, if I am what I think ." In W. Tschacher & C. Bergomi (Eds.), *The implications of embodiment (Cognition and Communication)* (pp. 89–125). Imprint Academic.

Friston, K. (2013a). Active inference and free energy. *Behavioral and Brain Sciences*, *36*(03), 212–213. doi:10.1017/S0140525X12002142

Friston, K. (2013b). Life as we know it. *Journal of the Royal Society, Interface* 10(86), 20130475. doi:10.1098/rsif.2013.0475

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, *2*(1-2), 56–78. doi:10.1002/hbm.460020107

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Front. Psychol.*, *3*, 151. doi:10.3389/fpsyg.2012.00151.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological Cybernetics*, *102*(3), 227–60. doi:10.1007/s00422-010-0364-z

Friston, K. J., & Price, C. J. (2001). Dynamic representations and generative models of brain function, *Brain Res. Bull.*, *54*(3), 275–285.

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., … Bestmann, S. (2012). Dopamine, affordance and active inference. *PLoS Computational Biology*, *8*(1), e1002327. doi:10.1371/journal.pcbi.1002327

Friston, K., Levin, M., Sengupta, B., & Pezzulo, G. (2015). Knowing one's place : a free-energy approach to pattern regulation *J. R. Soc. Interface* 12: 20141383. http://dx.doi.org/10.1098/rsif.2014.1383

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., FitzGerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience* 150217111908007. doi:10.1080/17588928.2015.1020053

Gershman, S. J., & Daw, N. D. (2011). Perception , Action and Utility : The Tangled Skein. In M. Rabinovich, M., K. Friston, & P. Varona (Eds.), *Principles of brain dynamics: Global state interactions* 293-312. Cambridge, MA: MIT Press.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Grau, C. (2005). *Philosophers Explore The Matrix*. Oxford University Press.

Gregory, R. (1980). Perceptions as hypotheses. *Phil. Trans. R. Soc. Lond., Series B, Biological Sciences*, *290*(1038), 181–197.

Haddock, A., & Macpherson, F. (Eds.). (2008). *Disjunctivism: Perception, action, and knowledge*. Oxford: Oxford University Press.

Haugeland, J. (2000). Mind Embodied and Embedded. In J. Haugeland, *Having Thought* (pp. 207–237).

Hempel, C. G. (1965). *Aspects of scientific explanation: and other essays in the philosophy of science*. Free Press.

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, *11*(10), 428–34. doi:10.1016/j.tics.2007.09.004

Hinton, G. E., & von Camp, D. (1993). Keeping neural networks simple by minimizing the description length of weights. *Proceedings of COLT-93*, 5–13.

Hinton, G. E., & Zemel, R. S. (1994). Autoencoders, minimum description length and Helmholtz free energy. In J. Cowan, G. Tesauro, & J. Alspector (Eds.), *Advances in neural information processing systems*, 6 . San Mateo, CA: Morgan Kaufmann.

Hipp, J. F., Engel, A. K., & Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron*, *69*, 387–396.

Hobson, J. A., & Friston, K. J. (2014). Consciousness, Dreams, and Inference: The Cartesian Theatre Revisited. *Journal of Consciousness Studies*, *21*(1-2), 6–32.

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

Hohwy, J. (2014). The self-evidencing brain. *Nous*, *00*, 1–27. doi:10.1111/nous.12062

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, *108*(3), 687–701.

Hon, G., & Rakover, S. S. (2013). *Explanation: Theoretical Approaches and Applications*. Springer Netherlands.

Horwitz, B. (2003). The elusive concept of brain connectivity. *Neuroimage*, *19*, 466–470.

Hoshi, E., & Tanji, J. (2007). Distinctions between dorsal and ventral premotor areas: anatomical connectivity and functional properties. *Current Opinion in Neurobiology*. doi:10.1016/j.conb.2007.02.003

Huang, Y., & Rao, R. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 580–593.

Hurley, S. (1998). *Consciousness in action*. Cambridge, MA: Harvard University Press

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321.

Kirsh, D., & Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, *18*(4), 513–549. doi:10.1207/s15516709cog1804_1

Kveraga, K., Ghuman, A., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, *65*, 145–168.

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: a paradigm of ecological optics. *Nature*. 293: 293–294.

MacKay, D. J. C. (1995). Free-energy minimization algorithm for decoding and cryptoanalysis. *Electron Lett.*, *31*, 445–447.

MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press.

Menary, R. (Ed.). (2010). *The extended mind*. Cambridge, MA: MIT Press.

Neal, R. M., & Hinton, G. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In M. I. Jordan (Ed.), *Learning in Graphical Models* (pp. 355–368). Dordrecht: Kluwer.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609. doi:10.1038/381607a0

Pezzulo, G., Rigoli, F., & Chersi, F. (2013). The mixed instrumental controller: using value of information to combine habitual choice and mental simulation. *Frontiers in Psychology*, *4*(March), 92. doi:10.3389/fpsyg.2013.00092

Pfeifer, R., & Bongard, J. (2006). *How the body shapes the way we think: A new view of intelligence*. Cambridge, MA: MIT Press.

Plato., & Jowett, B. (1941). *Plato's the republic*. New York: The Modern library.

Popper, K. R. (2002). *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge.

Powell, K. D., & Goldberg, M. E. (2000). Response of Neurons in the Lateral Intraparietal Area to a Distractor Flashed During the Delay Period of a Memory-Guided Saccade. *Journal of Neurophysiology*, *84*(1), 301–310.

Powers, William T. (2005). *Behavior: The control of perception.* New Canaan: Benchmark Publications

Rao, R., & Ballard, D. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79.

Riddoch, M. J., Humphreys, G. W., & Edwards, M. G. (2000). Neuropsychological evidence distinguishing object selection from action (effector) selection. *Cognitive Neuropsychology*, *17*, 547–562. doi:10.1080/02643290050110647

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy*, *101*(8), 389–428.

Rupert, R. (2009). *Cognitive systems and the extended mind.* New York: Oxford University Press.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–73. doi:10.1016/j.tics.2013.09.007

Smith, L., & Gasser, M. (2005). The development of embodied cognition: Six lesson from babies. *Artificial Life*, *11*(1), 13–30.

Sporns, O. (2010). *Networks of the brain*. Cambridge, MA: MIT Press.

Stephan, K. E., Harrison, L. M., Kiebel, S. J., David, O., Penny, W. D., & Friston, K. J. (2007). Dynamic causal models of neural system dynamics:current state and future extensions. *Journal of Biosciences*, *32*(1), 129–144. doi:10.1007/s12038-007-0012-5

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, *331*(6022), 1279–1285.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, *24*, 1–33.

Thompson, E., & Varela, F. J. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences*. doi:10.1016/S1364-6613(00)01750-2

Tresilian, J. R. (1999). Visually timed action: Time-out for "tau"? *Trends in Cognitive Sciences.*

Turvey, M. T., & Shaw, R. E. (1999). Ecological foundations of cognition: I. Symmetry and specificity of animal-environment systems. *Journal of Consciousness Studies*, *6*, 85–110.

Warren, W. (2006). The dynamics of action and perception. *Psychological Review* *113*(2), 358–389.

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

[1] Every learning system must start somewhere, and there is an important sense in which even arbitrary initial choices (such as numbers of levels, and numbers of units per level) can be seen as building in biases that affect subsequent learning. Biological systems come equipped with multiple extremely non-arbitrary initial structures including sensors, sensor positioning, wiring, and gross morphology. Such systems are already well on the road to knowing their own bodies and worlds. For important work on how minimal early biases can guide learning, see Elman (2005), Carey (2009).

[2] Examples include EM (expectation maximization – see Dempster et al (1977)) and various forms of gradient descent learning (see (MacKay, 2003)).

[3] Examples include Rao and Ballard (1999), Kemp et al (2007), Tenenbaum et al (2011).

[4] For an early example, see Rao and Ballard (1999). See also the important body of work by Karl Friston and collaborators summarized in Friston (2009) and (2010). For comprehensive reviews, see Hohwy (2013), Clark (2013), and (with a somewhat different focus) Hinton (2007) and Huang and Rao (2011).

[5] The term 'predictive processing' (PP) is used by Clark (2013), and 'prediction error minimizing' (PEM) by Hohwy (2013). In what follows, I shall stick with PP, but the reader may substitute these for one another at will.

[6] For reviews, see Kveraga et al (2007), Bubic et al (2010), Hohwy (2013), Clark (2013), Clark (2016))

[7] For some nice simulation studies displaying the power of such active interrogations, see Friston, Adams, Perrinet, and Breakspear (2012), who describe visual search as a form of hypothesis-testing experiment. As long as no hurdles (prediction error signals) are encountered, the current hypothesis remains in place. But that hypothesis is continually put the test. Here, as elsewhere, the epistemology of predictive processing is reminiscent of the falsificationist agenda associated with Karl Popper – see e.g. (Popper, 1963/2002).

[8] See Ballard (1991), Churchland et al (1994), Warren (2006), Anderson (2014) pp 163-172.

[9] Thus Hohwy et al (2008) note that "Terms like 'predictions' and 'hypotheses' sound rather intellectualist when it comes to basic perceptual inference. But at its heart the only processing aim of the system is simply to minimize prediction error or free energy, and indeed, the talk of hypotheses and predictions can be translated into such a less anthropomorphic framework [and] implemented using relatively simple neuronal infrastructures." (Hohwy et al. 2008, pp. 688–690)

[10] One might, in fact, deny that evil demon style manipulations actually deceive us. Instead, they might merely create an alternate substrate for the same old veridical knowledge about an external reality built of tables, chairs, baseball games and the like. For this kind of response, see Chalmers (2005). Alternatively, one might espouse a disjunctivist view of the contents of perceptual experience (for discussion, see the essays in Haddock and Macpherson 2008).

[11] This does bear in one way on the skeptical challenge. For when perception plays a non-reconstructive role, as Anderson (op cit p.185) notes "There is no need to posit any further epistemic mediators… to characterize the nature of [the] perception-action coupling."

[12] Typically, these rich inner models involved symbolic encodings that described states of affairs using complex language-like knowledge structures. Nothing like this is implied by PEM.

[13] For a thorough rehearsal of the positive arguments, see Clark (2008). For critiques, see Rupert (2004) (2009), (Adams & Aizawa (2008). For a rich sampling of the ongoing debate, see the essays in Menary (2010).

[14] Similarly, when we solve a mathematical puzzle with pen and paper, we do not need to represent all the details in our heads – instead, we rely on the properties of the external medium to bear some of the problem-solving load.

[15] Such considerations are not, in themselves, sufficient to establish the truth of the extended mind hypothesis (for such an argument, the reader is referred to Clark (2008)). But they do suggest that there is ample space for such a story within the PP/PEM vision of the contribution of the biological brain.

[16] Free-energy formulations originate in statistical physics and were introduced into the machine-learning literature in seminal treatments that include Hinton and von Camp (1993), Hinton and Zemel (1994), MacKay (1995), and Neal and Hinton (1998). Such formulations can arguably be used (e.g., Friston, 2010) to display the prediction error minimization strategy as *itself* a manifestation of a more fundamental mandate to minimize an information-theoretic isomorph of thermodynamic free energy in a system's exchanges with the environment.

[17] This deeper ambivalence in the notion of inference is nicely captured by Bruineberg and Rietveld who note that "within philosophy and cognitive science the notion of "inference" is traditionally understood in terms of arriving at a propositional statement based on some premises or observations. Within the Free Energy framework the notion of "inference" is much more minimal and does not involve any propositions: any dynamical system A coupled with another B can be said to "infer" the "hidden cause" of its "input" (the dynamics of B) when it reliably covaries with the dynamics of B and it is robust to the noise inherent in the coupling" Bruineberg and Rietveld (2014, p.7)

[18] In essence, I would opt here for a broadly pragmatist response. The goodness of our bedrock worldly understanding, I would argue, is simply constituted by the way it enables ongoing perception/action loops to succeed. Such a view leaves no room for what Hohwy (2014, p. 5) describes as a 'schism' between prediction-generating neural models and modeled states of the world. For (bracketing the special contexts set up by scientific practice – there is a whole other discussion to be had thereabouts) our neural models are not meant to be descriptions of an agent-independent world. Rather, they are recipes for engaging that world. They are recipes, moreover, that can rely heavily upon the reliable contributions of bodily and environmental factors and forces. This offers some levers for thinking about the 'extended mind', as we see in section 5.

[19] This implies – just as Hohwy and Friston insist – that any viable agent, in its many exchanges with the world, maximizes the sensory evidence for the model that it (in this extended sense) embodies. We already saw (section 3) that this should not be taken as evidence for the use of a fully reconstructive inner model. Similarly, we now see that the notion of self-evidencing does not impose a sensation-based veil that screens off the extra-neural world.

[20] Efficiency and the use of minimal models are implied by the free energy principle, insofar as securing behavioral success at minimal energetic cost is simply part and parcel of organismic resistance to the second law of thermodynamics.

[21] Such a view would be suggested if, for example, one were to identify the PEM story with earlier perception-driven treatments such as Gregory (1980).

[22] 'Effective connectivity' (see Aertsen et al., (1987), (Friston (1995), Horwitz (2003), Sporns (2010)) names 'the influence one neural system exerts over another' ((K. J. Friston, 1994) p. 57). It is to be distinguished from both structural and functional connectivity. 'Structural connectivity' names the gross pattern of physical linkages (the web of fibers and synapses) that allow neurons to interact across space and time. 'Functional connectivity' describes observed patterns of temporal correlation between neural events. The closely related notion of 'effective connectivity' reflects short-term patterns of causal influence between neural events, taking us beyond simple observations of undirected – and sometimes uninformative - correlation.

[23] For more on such transient ensembles, see Anderson, Richardson, and Chemero (2012), Anderson (2014).

[24] For a lovely exploration of these two roles for action, and their fluid emergence from a predictive processing architecture, see Friston, Rigoli, et al. (2015).