Embodied Prediction

Andy Clark

Versions of the "predictive brain" hypothesis rank among the most promising and the most conceptually challenging visions ever to emerge from computational and cognitive neuroscience. In this paper, I briefly introduce (section 1) the most radical and comprehensive of these visions—the account of "active inference", or "action-oriented predictive processing" (Clark 2013a), developed by Karl Friston and colleagues. In section 2, I isolate and discuss four of the framework's most provocative claims: (i) that the core flow of information is top-down, not bottom-up, with the forward flow of sensory information replaced by the forward flow of prediction error; (ii) that motor control is just more top-down sensory prediction; (iii) that efference copies, and distinct "controllers", can be replaced by top-down predictions; and (iv) that cost functions can fruitfully be replaced by predictions. Working together, these four claims offer a tantalizing glimpse of a new, integrated framework for understanding perception, action, embodiment, and the nature of human experience. I end (section 3) by sketching what may be the most important aspect of the emerging view: its ability to embed the use of fast and frugal solutions (as highlighted by much work in robotics and embodied cognition) within an over-arching scheme that includes more structured, knowledge-intensive strategies, combining these fluently and continuously as task and context dictate.

Keywords
Active inference | Embodied cognition | Motor control | Prediction | Prediction error

Author

Andy Clark
andy.clark@ed.ac.uk
University of Edinburgh
Edinburgh, United Kingdom

Commentator

Michael Madary
madary@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Editors

Thomas Metzinger
metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt
jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 Mind turned upside down?

PP (Predictive processing; for this terminology, see Clark 2013a) turns a traditional picture of perception on its head. According to that once-standard picture (Marr 1982), perceptual processing is dominated by the forward flow of information transduced from various sensory receptors. As information flows forward, a progressively richer picture of the real-world scene is constructed. The process of construction would involve the use of stored knowledge of various kinds, and the forward flow of information was subject to modulation and nuancing by top-down (mostly attentional) effects. But the basic picture remained one in which perception was fundamentally a process of "bottom-up feature detection". In Marr's theory of vision, detected intensities (arising from surface discontinuities and other factors) gave way to detected features such as blobs, edges, bars, "zero-crossings", and lines, which in turn gave way to detected surface orientations leading ultimately (though this step was always going to be problematic) to a three-dimensional model of the visual scene. Early perception is here seen as building towards a complex world model by a feedforward process of evidence accumulation. Traditional perceptual neuroscience followed suit, with visual cortex (the most-studied example) being "traditionally viewed as a hierarchy of neural feature detectors, with neural population responses being driven by bottom-up stimulus features" (Egner et al. 2010, p. 16601). This was a view of the perceiving brain as passive and stimulus-driven, taking energetic inputs

Clark, A. (2015). Embodied Prediction.
In T. Metzinger & J. M. Windt (Eds). *Open MIND:* 7(T). Frankfurt am Main: MIND Group. doi: 10.15502/9783958570115

from the senses and turning them into a coherent percept by a kind of step-wise build-up moving from the simplest features to the more complex: from simple intensities up to lines and edges and on to complex meaningful shapes, accumulating structure and complexity along the way in a kind of Lego-block fashion.

Such views may be contrasted with the increasingly active views that have been pursued over the past several decades of neuroscientific and computational research. These views (Ballard 1991; Churchland et al. 1994; Ballard et al. 1997) stress the active search for task-relevant information just-in-time for use. In addition, huge industries of work on intrinsic neural activity, the "resting state" and the "default mode" (for a review, see Raichle & Snyder 2007) have drawn our attention to the ceaseless buzz of neural activity that takes place even in the absence of ongoing task-specific stimulation, suggesting that much of the brain's work and activity is in some way ongoing and endogenously generated.

Predictive processing plausibly represents the last and most radical step in this retreat from the passive, input-dominated view of the flow of neural processing. According to this emerging class of models, naturally intelligent systems (humans and other animals) do not passively await sensory stimulation. Instead, they are constantly active, trying to predict the streams of sensory stimulation before they arrive. Before an "input" arrives on the scene, these pro-active cognitive systems are already busy predicting its most probable shape and implications. Systems like this are already (and almost constantly) poised to act, and all they need to process are any sensed deviations from the predicted state. It is these calculated deviations from predicted states (known as *prediction errors*) that thus bear much of the information-processing burden, informing us of what is salient and newsworthy within the dense sensory barrage. The extensive use of top-down probabilistic prediction here provides an effective means of avoiding the kinds of "representational bottleneck" feared by early opponents (e.g., Brooks 1991) of representation-heavy—but feed-forward dominated—forms of processing. Instead, the downward flow of prediction now does most of the computational "heavy-lifting", allowing moment-by-moment processing to focus only on the newsworthy departures signified by salient (that is, high-precision—see section 3) prediction errors. Such economy and preparedness is biologically attractive, and neatly sidesteps the many processing bottlenecks associated with more passive models of the flow of information.

Action itself (more on this shortly) then needs to be reconceived. Action is not so much a response to an input as a neat and efficient way of selecting the next "input", and thereby driving a rolling cycle. These hyperactive systems are constantly predicting their own upcoming states, and actively moving so as to bring some of them into being. We thus act so as to bring forth the evolving streams of sensory information that keep us viable (keeping us fed, warm, and watered) and that serve our increasingly recondite ends. PP thus implements a comprehensive reversal of the traditional (bottom-up, forward-flowing) schema. The largest contributor to ongoing neural response, if PP is correct, is the ceaseless anticipatory buzz of downwards-flowing neural prediction that drives both perception and action. Incoming sensory information is just one further factor perturbing those restless pro-active seas. Within those seas, percepts and actions emerge via a recurrent cascade of sub-personal predictions forged (see below) from unconscious expectations spanning multiple spatial and temporal scales.

Conceptually, this implies a striking reversal, in that the driving sensory signal is really just providing corrective feedback on the emerging top-down predictions.[1] As ever-active prediction engines, these kinds of minds are not, fundamentally, in the business of solving puzzles given to them as inputs. Rather, they are in the business of keeping us one step ahead of the game, poised to act and actively eliciting the sensory flows that keep us viable and fulfilled. If this is on track, then just about every aspect of the passive forward-flowing model is false. We are not passive cognitive couch potatoes so

1 For this observation, see Friston (2005), p. 825, and the discussion in Hohwy (2013).

much as proactive predictavores, forever trying to stay one step ahead of the incoming waves of sensory stimulation.

2 Radical predictive processing

Such models involve a number of quite radical claims. In the present treatment, I propose focusing upon just four:

1. The core flow of information is top-down, not bottom-up, and the forward flow of sensory information is replaced by the forward flow of prediction error.
2. Motor control is just more top-down sensory prediction.
3. Efference copies, and distinct "controllers" (inverse models) are replaced by top-down predictions.
4. Cost functions are absorbed into predictions.

One thing I shan't try to do here is rehearse the empirical evidence for the framework. That evidence (which is substantial but importantly incomplete) is rehearsed in Clark (2013a) and Hohwy (2013, this collection). For a recent attempt to specify a neural implementation, see Bastos et al. (2012). I now look at each of these points in turn:

2.1 The core flow of information is top-down, not bottom-up, and the forward flow of sensory information is replaced by the forward flow of prediction error

This is the heart and soul of the radical vision. Incoming sensory information, if PP is correct, is constantly met with a cascade of top-down prediction, whose job is to predict the incoming signal across multiple temporal and spatial scales.

To see how this works in practice, consider a seminal proof-of-concept by Rao & Ballard (1999). In this work, prediction-based learning targets image patches drawn from natural scenes using a multi-layer artificial neural network. The network had no pre-set task apart from that of using the downwards connections to match input samples with successful predictions. Instead, visual signals were processed via a hierarchical system in which each level tried (in the way just sketched) to predict activity at the level below it using recurrent (feedback) connections. If the feedback successfully predicted the lower-level activity, no further action was required. Failures to predict enabled tuning and revision of the model (initially, just a random set of connection weights) generating the predictions, thus slowly delivering knowledge of the regularities governing the domain. In this architecture, forward connections between levels carried only the "residual errors" (Rao & Ballard 1999, p. 79) between top-down predictions and actual lower level activity, while backward or recurrent connections carried the predictions themselves.

After training, the network developed a nested structure of units with simple-cell-like receptive fields and captured a variety of important, empirically-observed effects. One such effect was "end-stopping". This is a "non-classical receptive field" effect in which a neuron responds strongly to a short line falling within its classical receptive field but (surprisingly) shows diminishing response as the line gets longer. Such effects (and with them, a whole panoply of "context effects") emerge naturally from the use of hierarchical predictive processing. The response tails off as the line gets longer, because longer lines and edges were the statistical norm in the natural scenes to which the network was exposed in training. After training, longer lines are thus what is first predicted (and fed back, as a hypothesis) by the level-two network. The strong firing of some level-one "edge cells", when they are driven by shorter lines, thus reflects not successful feature detection by those cells but rather error or mismatch, since the short segment was not initially predicted by the higher-level network. This example neatly illustrates the dangers of thinking in terms of a simple cumulative flow of feature-detection, and also shows the advantages of re-thinking the flow of processing as a mixture of top-down prediction and bottom-up error correction.[2] In ad-

2 This does not mean that there are no cells in v1 or elsewhere whose responses match the classical profile. PP claims that each neural area

dition it highlights the way these learning routines latch on to the world in a manner specified by the training data. End-stopped cells are simply a response to the structure of the natural scenes used in training, and reflect the typical length of the lines and edges in these natural scenes. In a very different world (such as the underwater world of some sea-creatures) such cells would learn very different responses.

These were early and relatively low-level results, but the predictive processing model itself has proven rich and (as we shall see) widely applicable. It assumes only that the environment generates sensory signals by means of nested interacting causes and that the task of the perceptual system is to invert this structure by learning and applying a structured internal model—so as to predict the unfolding sensory stream. Routines of this kind have recently been successfully applied in many domains, including speech perception, reading, and recognizing the actions of oneself and of other agents (see Poeppel & Monahan 2011; Price & Devlin 2011; Friston et al. 2011). This is not surprising, since the underlying rationale is quite general. If you want to predict the way some set of sensory signals will change and evolve over time, a good thing to do is to learn how those sensory signals are determined by interacting external causes. And a good way to learn about those interacting causes is to try to predict how the sensory signal will change and evolve over time.

Now try to imagine this this on a very grand scale. To predict the visually presented scene, the system must learn about edges, blobs, line segments, shapes, forms, and (ultimately) objects. To predict text, it must learn about interacting "hidden" causes in the linguistic domain: causes such as sentences, words, and letters. To predict all of our rich multimodal plays of sensory data, across many scales of space and time, it must learn about interacting hidden causes such as tables, chairs, cats, faces, people, hurricanes, football games, goals, meanings, and intentions. The structured world of human experience, if this is correct, comes into view only when all manner of top-down predictions meet (and "explain away") the incoming waves of sensory information. What propagates forwards (through the brain, away from the sensory peripheries) is then only the mismatches, at every level, with predicted activity.

This makes functional sense. Given that the brain is ever-active, busily predicting its own states at many levels, all that matters (that is, all that is newsworthy, and thus ought to drive further processing) are the incoming surprises: unexpected deviations from what is predicted. Such deviations result in prediction errors reflecting residual differences, at every level and stage of processing, between the actual current signal and the predicted one. These error signals are used to refine the prediction until the sensory signal is best accommodated.

Prediction error thus "carries the news", and is pretty much the hero (or anti-hero) of this whole family of models. So much so, that it is sometimes said that:

> In predictive coding schemes, sensory data are replaced by prediction error, because that is the only sensory information that has yet to be explained. (Feldman & Friston 2010, p. 2)

We can now savor the radicalism. Where traditional, feed-forward-based views see a progressive (though top-down modulated) flow of complex feature-detection, the new view depicts a progressive, complex flow of feature prediction. The top-down flow is not mere attentional modulation. It is the core flow of structured content itself. The forward-flowing signal, by contrast, has now morphed into a stream of residual error. I want to suggest, however, that we treat this apparently radical inversion with some caution. There are two reasons for this—one conceptual, and one empirical.

The first (conceptual) reason for caution is that the "error signal" in a trained-up predictive coding scheme is highly informative. Prediction error signals carry detailed information

contains two kinds of cell, or at least supports two functionally distinct response profiles, such that one functionality encodes error and the other current best-guess content. This means that there can indeed be (as single cell recordings amply demonstrate) recognition cells in each area, along with the classical response profiles. For more on this important topic, see Clark (2013a).

about the mismatched content itself. Prediction errors are thus as structured and nuanced in their implications as the model-based predictions relative to which they are computed. This means that, in a very real sense, the prediction error signal is not a mere proxy for incoming sensory information – it *is* sensory information. Thus, suppose you and I play a game in which I (the "higher, predicting level") try to describe to you (the "lower level") the scene in front of your eyes. I can't see the scene directly, but you can. I do, however, believe that you are in some specific room (the living room in my house, say) that I have seen in the past. Recalling that room as best I can, I say to you "there's a vase of yellow flowers on a table in front of you". The game then continues like this. If you are silent, I take that as your agreeing to my description. But if I get anything that matters wrong, you must tell me what I got wrong. You might say "the flowers are yellow". You thus provide an error signal that invites me to try again in a rather specific fashion—that is, to try again with respect to the colour of the flowers in the vase. The next most probable colour, I conjecture, is red. I now describe the scene in the same way but with red flowers. Silence. We have settled into a mutually agreeable description.[3]

The point to note is that your "error signal" carried some quite specific information. In the pragmatic context of your silence regarding all other matters, the content might be glossed as "there is indeed a vase of flowers on the table in front of me but they are not yellow". This is a pretty rich message. Indeed, it does not (content-wise) seem different in kind to the downward-flowing predictions themselves. Prediction error signals are thus richly informative, and as such (I would argue) not radically different to sensory information itself. This is unsurprising, since mathematically (as Karl Friston has pointed out[4]) sensory information and prediction error are informationally identical, except that the latter are centred on the predictions. To see this, reflect on the fact that prediction error is just the original information minus the prediction. It follows that the original information is given by the prediction error plus the prediction. Prediction error is simply error relative to some specific prediction and as such it flags the sensory information that is as yet unexplained. The forward flow of prediction error thus constitutes a *forward flow of sensory information relative to specific predictions*.

There is more to the story at this point, since the (complex, non-linear) ways in which downward-flowing predictions interact are importantly different to the (simple, linear) effects of upward-flowing error signals. Non-linearities characterize the multi-level construction of the predictions, which do the "heavy lifting", while the prediction error signals are free to behave additively (since all the complex webs of linkage are already in place). But the bottom line is that prediction error does not replace sensory information in any mysterious or conceptually challenging fashion, since prediction error is nothing other than that sensory information that has yet to be explained.

The second (empirical) reason for caution is that it is, in any case, only one specific implementation of the predictive brain story depicts the forward-flow as consisting solely of prediction error. An alternative implementation (due to Spratling 2008 and 2010—and see discussion in Spratling 2013) implements the same key principles using a different flow of prediction and error, and described by a variant mathematical framework. This illustrates the urgent need to explore multiple variant architectures for prediction error minimization. In fact, the PP schema occupies just one point in a large and complex space of probabilistic generative-

3 To complete the image using this parlour game, we'd need to add a little more structure to reflect the hierarchical nature of the message-passing scheme. We might thus imagine many even-higher-level "prediction agents" working together to predict which room (house, world, etc.) the layers below are currently responding to. Should sufficient prediction error signals accrue, this ensemble might abandon the hypothesis that signals are coming in from the living room, suggesting instead that they are from the boudoir, or the attic. In this grander version (which recalls the "mixtures of experts" model in machine learning—see Jordan & Jacobs 1994)—there are teams (and teams of teams) of specialist prediction agents, all trying (guided top-down by the other prediction agents, and bottom-up by prediction errors from the level below) to decide which specialists should handle the current sensory barrage. Each higher-level "prediction agent", in this multi-level version, treats activity at the level below as sensory information, to be explained by the discovery of apt top-down predictions.

4 Personal communication.

model-based approaches, and there are many possible architectures and possible ways of combining top-down predictions and bottom-up sensory information in this general vicinity. These include foundational work by Hinton and colleagues on deep belief networks (Hinton & Salakhutdinov 2006; Hinton et al. 2006), work that shares a core emphasis on the use of prediction and probabilistic multi-level generative models, as well as recent work combining connectionist principles with Bayesian angles (see McClelland 2013 and Zorzi et al. 2013). Meanwhile, roboticists such as Tani (2007), Saegusa et al. (2008), Park et al. (2012), Pezzulo (2008), and Mohan et al. (2010) explore the use of a variety of prediction-based learning routines as a means of grounding higher cognitive functions in the solid bedrock of sensorimotor engagements with the world. Only by considering the full space of possible prediction-and-generative-model-based architectures and strategies can we start to ask truly pointed experimental questions about the brain and about biological organisms; questions that might one day favor one of these models (or, more likely, one coherent sub-set of models[5]) over the rest, or else may reveal deep faults and failings among their substantial common foundations.

2.2 Motor control is just more top-down sensory prediction

I shall, however, continue to concentrate upon the specific explanatory schema implied by PP, as this represents (it seems to me) the most comprehensive and neuroscientifically well-grounded vision of the predictive mind currently available. What makes PP especially interesting —and conceptually challenging—is the seamless integration of perception and action achieved under the rubric of "active inference".

To understand this, consider the motor system. The motor system (like the visual cortex) displays a complex hierarchical structure.[6] Such a structure allows complex behaviors to be specified, at higher levels, in compact ways, the implications of which can be progressively unpacked at the lower levels. The traditional way of conceptualizing the difference, however, is that in the case of motor control we imagine a downwards flow of information, whereas in the case of the visual cortex we imagine an upwards flow. Descending pathways in the motor cortex, this traditional picture suggests, should correspond functionally to ascending pathways in the visual cortex. This is not, however, the case. Within the motor cortex the downwards connections (descending projections) are "anatomically and physiologically more like backwards connections in the visual cortex than the corresponding forward connections" (Adams et al. 2013, p. 1).

This is suggestive. Where we might have imagined the functional anatomy of a hierarchical motor system to be some kind of inverted image of that of the perceptual system, instead the two seem fundamentally alike.[7] The explanation, PP suggests, is that the downwards connections, in both cases, take care of essentially the same kind of business—namely the business of predicting sensory stimulation. Predictive processing models subvert, we saw, the traditional picture with respect to perception. In PP, compact higher-level encodings are part of an apparatus that tries to predict the play of energy across the sensory surfaces. The same story applies, recent extensions (see below) of PP suggest, to the motor case. The difference is that motor control is, in a certain sense, subjunctive. It involves predicting the non-actual sensory trajectories that *would* ensue *were* we performing some desired action. Reducing prediction er-

5 One such subset is, of course, the set of hierarchical dynamic models (see Friston 2008).

6 The appeal to hierarchical structure in PP, it should be noted, is substantially different to that familiar from treatments such as Felleman & Van Essen (1991). Although I cannot argue for this here (for more on this see Clark 2013b; in press) the PP hierarchy is fluid in that the information-flows it supports are reconfigurable moment-by-moment (by, for example, changing be and theta band oscillations —see Bastos et al. 2015). In addition, PP dispenses entirely with the traditional idea (nicely reviewed, and roundly rejected, in Churchland et al. 1994) that earlier levels must complete their tasks before passing information "up" the hierarchy. The upshot is that the PP models are much closer to dynamical systems accounts than to traditional, feed forward, hierarchical ones.

7 For the full story, see Adams et al. (2013). In short: "[t]he descending projections from motor cortex share many features with top-down or backward connections in visual cortex; for example, corticospinal projections originate in infragranular layers, are highly divergent and (along with descending cortico-cortical projections) target cells expressing NMDA receptors" (Adams et al. 2013, p. 1).

rors calculated against these non-actual states then serves (in ways we are about to explore) to make them actual. We predict the sensory consequences of our own action and this brings the actions about.

The upshot is that the downwards connections, in both the motor and the sensory cortex, carry complex predictions, and the upwards connections carry prediction errors. This explains the otherwise "paradoxical" (Shipp et al. 2013, p. 1) fact that the functional circuitry of the motor cortex does not seem to be inverted with respect to that of the sensory cortex. Instead, the very distinction between the motor and the sensory cortex is now eroded—both are in the business of top-down prediction, though the kind of thing they predict is (of course) different. The motor cortex here emerges, ultimately, as a multimodal sensorimotor area issuing predictions in both proprioceptive and other modalities.

In this way, PP models have been extended (under the umbrella of "active inference"—see Friston 2009; Friston et al. 2011) to include the control of action. This is accomplished by predicting the flow of sensation (especially that of proprioception) that would occur were some target action to be performed. The resulting cascade of prediction error is then quashed by moving the bodily plant so as to bring the action about. Action thus results from our own predictions concerning the flow of sensation—a version of the "ideomotor" theory of James (1890) and Lotze (1852), according to which the very idea of moving, when unimpeded by other factors, is what brings the moving about. The resulting schema is one in which:

> The perceptual and motor systems should not be regarded as separate but instead as a single active inference machine that tries to predict its sensory input in all domains: visual, auditory, somatosensory, interoceptive and, in the case of the motor system, proprioceptive. (Adams et al. 2013, p. 4)

In the case of motor behaviors, the key driving predictions, Friston and colleagues suggest, are predictions of the proprioceptive patterns[8] that would ensue were the action to be performed (see Friston et al. 2010). To make an action come about, the motor plant responds so as to cancel out proprioceptive prediction errors. In this way, predictions of the unfolding proprioceptive patterns that would be associated with the performance of some action serve to bring that action about. Proprioceptive predictions directly elicit motor actions (so traditional motor commands are simply replaced by those proprioceptive predictions).

This erases any fundamental computational line between perception and the control of action. There remains, to be sure, an obvious (and important) difference in direction of fit. Perception here matches neural hypotheses to sensory inputs, and involves "predicting the present"; while action brings unfolding proprioceptive inputs into line with neural predictions. The difference, as Elizabeth Anscombe (1957) famously remarked,[9] is akin to that between consulting a shopping list to select which items to purchase (thus letting the list determine the contents of the shopping basket) and listing some actually purchased items (thus letting the contents of the shopping basket determine the list). But despite this difference in direction of fit, the underlying form of the neural computations is now revealed to be the same. Indeed, the main difference between the motor and the visual cortex, on this account, lies more in what kind of thing (for example, the proprioceptive consequences of a trajectory of motion) is predicted, rather than in how it is predicted. The upshot is that:

> The primary motor cortex is no more or less a motor cortical area than striate (visual) cortex. The only difference

8 Proprioception is the "inner" sense that informs us about the relative locations of our bodily parts and the forces and efforts that are being applied. It is to be distinguished from exteroceptive (i.e., standard perceptual) channels such as vision and audition, and from interoceptive channels informing us of hunger, thirst, and states of the viscera. Predictions concerning the latter may (see Seth 2013 and Pezzulo 2014) play a large role in the construction of feelings and emotions.

9 Anscombe's target was the distinction between desire and belief, but her observations about direction of fit generalize (as Shea 2013 notes) to the case of actions, here conceived as the motoric outcomes of certain forms of desire.

> between the motor cortex and visual cortex is that one predicts retinotopic input while the other predicts proprioceptive input from the motor plant. (Friston et al. 2011, p. 138)

Perception and action here follow the same basic logic and are implemented using the same computational strategy. In each case, the systemic imperative remains the same: the reduction of ongoing prediction error. This view has two rather radical consequences, to which we shall now turn.

2.3 Efference copies and distinct "controllers" are replaced by top-down predictions

A long tradition in the study of motor control invokes a "forward model" of the likely sensory consequences of our own motor commands. In this work, a copy of the motor command (known as the "efference copy"; Von Holst 1954) is processed using the forward model. This model captures (or "emulates"—see Grush 2004) the relevant biodynamics of the motor plant, enabling (for example) a rapid prediction of the likely feedback from the sensory peripheries. It does this by encoding the relationship between motor commands and predicted sensory outcomes. The motor command is thus captured using the efference copy which, fed to the forward model, yields a prediction of the sensory outcome (this is sometimes called the "corollary discharge"). Comparisons between the actual and the predicted sensory input are thus enabled.

But motor control, in the leading versions of this kind of account, requires in addition the development and use of a so-called "inverse model" (see e.g., Kawato 1999; Franklin & Wolpert 2011). Where the forward model maps current motor commands in order to predicted sensory effects, the inverse model (also known as a controller) "performs the opposite transformation [...] determining the motor command required to achieve some desired outcome" (Wolpert et al. 2003, p. 595). Learning and deploying an inverse model appropriate to some task is, however, generally much more demanding than learning the forward model, and requires solving a complex mapping problem (linking the desired end-state to a nested cascade of non-linearly interacting motor commands) while effecting transformations between varying co-ordinate schemes (e.g., visual to muscular or proprioceptive—see e.g., Wolpert et al. 2003, pp. 594–596).

PP (the full "action-inclusive" version just described) shares many key insights with this work. They have common a core emphasis on the prediction-based learning of a forward (generative) model, which is able to anticipate the sensory consequences of action. But active inference, as defended by Friston and others—see e.g., Friston (2011); Friston et al. (2012)—dispenses with the inverse model or controller, and along with it the need for efference copy of the motor command. To see how this works, consider that action is here reconceived as a direct consequence of predictions (spanning multiple temporal and spatial scales) about trajectories of motion. Of special importance here are predictions about proprioceptive consequences that implicitly minimize various energetic costs. Subject to the full cascade of hierarchical top-down processing, a simple motor command now unfolds into a complex set of predictions concerning both proprioceptive and exteroceptive effects. The proprioceptive predictions then drive behavior, causing us to sample the world in the ways that the current winning hypothesis dictates.[10]

Such predictions can be couched, at the higher levels, in terms of desired states or trajectories specified using extrinsic (world-centered, limb-centered) co-ordinates. This is possible because the required translation into intrinsic (muscle-based) co-ordinates is then devolved to what are essentially classical reflex arcs set up to quash priorioceptive prediction errors. Thus:

> if motor neurons are wired to suppress proprioceptive prediction errors in the dorsal horn of the spinal cord, they effect-

10 For a simulation-based demonstration of the overall shape of the PP account, see Friston et al. (2012). These simulations, as the authors note, turn out to implement the kind of "active vision" account put forward in Wurtz et al. (2011).

> ively implement an inverse model, mapping from desired sensory consequences to causes in intrinsic (muscle-based) coordinates. In this simplification of conventional schemes, descending motor commands become topdown predictions of proprioceptive sensations conveyed by primary and secondary sensory afferents. (Friston 2011, p. 491)

The need (prominent in approaches such as Kawato 1999; Wolpert et al. 2003; and Franklin & Wolpert 2011) for a distinct inverse model/optimal control calculation has now disappeared. In its place we find a more complex forward model mapping prior beliefs about desired trajectories to sensory consequences, some of which (the "bottom level" prorioceptive ones) are automatically fulfilled.

The need for efference copy has also disappeared. This is because descending signals are already (just as in the perceptual case) in the business of predicting sensory (both proprioceptive and exteroceptive) consequences. By contrast, so-called "corollary discharge" (encoding predicted sensory outcomes) is now endemic and pervades the downwards cascade, since:

> [...] every backward connection in the brain (that conveys topdown predictions) can be regarded as corollary discharge, reporting the predictions of some sensorimotor construct. (Friston 2011, p. 492)

This proposal may, on first encounter, strike the reader as quite implausible and indeed too radical. Isn't an account of the functional significance and neurophysiological reality of efference copy one of the major success stories of contemporary cognitive and computational neursocience? In fact, most (perhaps all) of the evidence often assumed to favour that account is, on closer examination, simply evidence of the pervasive and crucial role of forward models and corollary discharge—it is evidence, that is to say, for just those parts of the traditional story that are preserved (and made even more central) by PP. For example, Sommer & (Wurtz's influential (2008) review paper makes very little mention of efference copy as such, but makes widespread use of the more general concept of corollary discharge—though as those authors note, the two terms are often used interchangeably in the literature. A more recent paper, Wurtz et al. (2011), mentions efference copy only once, and does so only to merge it with discussions of corollary discharge (which then occur 114 times in the text). Similarly, there is ample reason to believe that the cerebellum plays a special role here, and that that role involves making or optimizing perceptual predictions about upcoming sensory events (Bastian 2006; Roth et al. 2013). But such a role is, of course, entirely consistent with the PP picture. This shows, I suggest, that it is the general concept of forward models (as used by e.g., Miall & Wolpert 1996) and corollary discharge, rather than the more specific concept of efference copy as we defined it above, that enjoys the clearest support from both experimental and cognitive neuroscience.

Efference copy figures prominently, of course, in one particular set of computational proposals. These proposals concern (in essence) the positioning of forward models and corollary discharges within a putative larger cognitive architecture involving multiple paired forward and inverse models. In these "paired forward inverse model" architectures (see e.g., Wolpert & Kawato 1998; Haruno et al. 2003) motor commands are copied to a stack of separate forward models that are used to predict the sensory consequences of actions. But acquiring and deploying such an architecture, as even its strongest advocates concede, poses a variety of extremely hard computational challenges (see Franklin & Wolpert 2011). The PP alternative neatly sidesteps many of these problems—as we shall see in section 2.4. The heavy lifting that is usually done by traditional efference copy, inverse models, and optimal controllers is now shifted to the acquisition and use of the predictive (generative) model—i.e., the right set of prior probabilistic "beliefs". This is potentially advantageous if (but only if) we can reasonably assume that these beliefs "emerge naturally as top-down or empirical priors during hierarchical perceptual inference" (Friston 2011, p. 492).

The deeper reason that efference copy may be said to have disappeared in PP is thus that the whole (problematic) structure of paired forward and inverse models is absent. It is not needed, because some of the predicted sensory consequences (the predicted proprioceptive trajectories) act as motor commands already. As a result, there are no distinct motor commands to copy, and (obviously) no efference copies as such. But one could equally well describe the forward-model-based predictions of proprioceptive trajectories as "minimal motor commands": motor commands that operate (in essence) by specifying results rather than by exerting fine-grained limb and joint control. These minimal motor commands (proprioceptive predictions) clearly influence the even wider range of predictions concerning the exteroceptive sensory consequences of upcoming actions. The core functionality that is normally attributed to the action of efference copy is thus preserved in PP, as is the forward-model-based explanation of core phenomena, such as the finessing of time-delays (Bastian 2006) and the stability of the visual world despite eye-movements (Sommer & Wurtz 2006; 2008).

2.4 Cost functions are absorbed by predictions.

Active inference also sidesteps the need for explicit cost or value functions as a means of selecting and sculpting motor response. It does this (Friston 2011; Friston et al. 2012) by, in essence, building these in to the generative model whose probabilistic predictions combine with sensory inputs in order to yield behaviors. Simple examples of cost or value functions (that might be applied to sculpt and select motor behaviors) include minimizing "jerk" (the rate of change of acceleration of a limb during some behavior) and minimizing rate of change of torque (for these examples see Flash & Hogan 1985 and Uno et al. 1989 respectively). Recent work on "optimal feedback control" minimizes more complex "mixed cost functions" that address not just bodily dynamics but also systemic noise and the required accuracy of outcomes (see Todorov 2004; Todorov & Jordan 2002). Such cost functions (as Friston 2011, p. 496 observes) resolve the many-one mapping problem that afflicts classical approaches to motor control. There are many ways of using one's body to achieve a certain goal, but the action system has to choose one way from the many available. Such devices are not, however, needed within the framework on offer, since:

> In active inference, these problems are resolved by prior beliefs about the trajectory (that may include minimal jerk) that uniquely determine the (intrinsic) consequences of (extrinsic) movements. (Friston 2011, p. 496)

Simple cost functions are thus folded into the expectations that determine trajectories of motion. But the story does not stop there. For the very same strategy applies to the notion of desired consequences and rewards at all levels. Thus we read that:

> Crucially, active inference does not invoke any "desired consequences". It rests only on experience-dependent learning and inference: experience induces prior expectations, which guide perceptual inference and action. (Friston et al. 2011, p. 157)

Notice that there is no *overall* computational advantage to be gained by this reallocation of duties. Indeed, Friston himself is clear that:

> [...] there is no free lunch when replacing cost functions with prior beliefs [since] it is well-known [Littman et al. (2001)] that the computational complexity of a problem is not reduced when formulating it as an inference problem. (2011, p. 492)

Nonetheless, it may well be that this reallocation (in which cost functions are treated as priors) has conceptually and strategically important consequences. It is easy, for example, to specify whole paths or trajectories using prior beliefs about (you guessed it) paths and trajectories! Scalar reward functions, by contrast, specify points or peaks. The upshot is that everything

that can be specified by a cost function can be specified by some prior over trajectories, but not vice versa.

Related concerns have led many working roboticists to argue that explicit cost-function-based solutions are inflexible and biologically unrealistic, and should be replaced by approaches that entrain actions in ways that implicitly exploit the complex attractor dynamics of embodied agents (see e.g., Thelen & Smith 1994; Mohan & Morasso 2011; Feldman 2009). One way to imagine this broad class of solutions (for a longer discussion, see Clark 2008, Ch. 1) is by thinking of the way you might control a wooden marionette simply by moving the strings attached to specific body parts. In such cases:

> The distribution of motion among the joints is the "passive" consequence of the [...] forces applied to the end-effectors and the "compliance" of different joints. (Mohan & Morasso 2011, p. 5)

Solutions such as these, which make maximal use of learnt or inbuilt "synergies" and the complex bio-mechanics of the bodily plant, can be very fluently implemented (see Friston 2011; Yamashita & Tani 2008) using the resources of active inference and (attractor-based) generative models. For example, Namikawa et al. (2011) show how a generative model with multi-timescale dynamics enables a fluent and decomposable (see also Namikawa & Tani 2010) set of motor behaviors. In these simulations:

> Action per se, was a result of movements that conformed to the proprioceptive predictions of [...] joint angles [and] perception and action were both trying to minimize prediction errors throughout the hierarchy, where movement minimized the prediction errors at the level of proprioceptive sensations. (Namikawa et al. 2011, p. 4)

Another example (which we briefly encountered in the previous section) is the use of downward-flowing prediction to side-step the need to transform desired movement trajectories from extrinsic (task-centered) to intrinsic (e.g., muscle-centered) co-ordinates: an "inverse problem" that is said to be both complex and ill-posed (Feldman 2009; Adams et al. 2013, p. 8). In active inference the prior beliefs that guide motor action already map predictions couched (at high levels) in extrinsic frames of reference onto proprioceptive effects defined over muscles and effectors, simply as part and parcel of ordinary online control.

By re-conceiving cost functions as implicit in bodies of expectations concerning trajectories of motion, PP-style solutions sidestep the need to solve difficult (often intractable) optimality equations during online processing (see Friston 2011; Mohan & Morasso 2011) and—courtesy of the complex generative model—fluidly accommodate signaling delays, sensory noise, and the many-one mapping between goals and motor programs. Alternatives requiring the distinct and explicit computation of costs and values thus arguably make unrealistic demands on online processing, fail to exploit the helpful characteristics of the physical system, and lack biologically plausible means of implementation.

These various advantages come, however, at a price. For the full PP story now shifts much of the burden onto the acquisition of those prior "beliefs"—the multi-level, multi-modal webs of probabilistic expectation that together drive perception and action. This may turn out to be a better trade than it at first appears, since (see Clark in in press) PP describes a biologically plausible architecture that is just about maximally well-suited to installing the requisite suites of prediction, through embodied interactions with the training environments that we encounter, perturb, and—at several slower timescales—actively construct.

3 Putting predictive processing, body, and world together again

An important feature of the full PP account (see Friston 2009; Hohwy 2013; Clark in press) is that the impact of specific prediction error signals can be systematically varied according to their estimated certainty or "precision". The precision of a specific prediction error is

its inverse variance—the size (if you like) of its error bars. Precision estimation thus has a kind of meta-representational feel, since we are, in effect, estimating the uncertainty of our own representations of the world. These ongoing (task and context-varying) estimates alter the weighting (the gain or volume, to use the standard auditory analogy) on select prediction error units, so as to increase the impact of task-relevant, reliable information. One key effect of this is to allow the brain to vary the balance between sensory inputs and prior expectations at different levels (see Friston 2009, p. 299) in ways sensitive to task and context.[11] High-precision prediction errors have greater gain, and thus play a larger role in driving processing and response. More generally, variable precision-weighting may be seen as the PP mechanism for implementing a wide range of attentional effects (see Feldman & Friston 2010).

Subtle applications of this strategy, as we shall shortly see, allow PP to nest simple ("quick and dirty") solutions within the larger context of a fluid, re-configurable inner economy; an economy in which rich, knowledge-based strategies and fast, frugal solutions are now merely different expressions of a unified underlying web of processing. Within that web, changing ensembles of inner resources are repeatedly recruited, forming and dissolving in ways determined by external context, current needs, and (importantly) by flexible precision-weighting reflecting ongoing estimations of our own uncertainty. This process of inner recruitment is itself constantly modulated, courtesy of the complex circular causal dance of sensorimotor engagement, by the evolving state of the external environment. In this way (as I shall now argue) many key insights from work on embodiment and situated, world-exploiting action may be comfortably accommodated within the emerging PP framework.

11 Malfunctions of this precision-weighting apparatus have recently been implicated in a number of fascinating proposals concerning the origins and persistence of various forms of mental disturbance, including the emergence of delusions and hallucinations in schizophrenia, "functional motor and sensory symptoms", Parkinson's disease, and autism—see Fletcher & Frith (2009), Frith & Friston (2012), Adams et al. (2012), Brown et al. (2013), Edwards et al. (2012), and Pellicano & Burr (2012).

3.1 Nesting simplicity within complexity

Consider the well-known "outfielder's problem": running to catch a fly ball in baseball. Giving perception its standard role, we might assume that the job of the visual system is to transduce information about the current position of the ball so as to allow a distinct "reasoning system" to project its future trajectory. Nature, however, seems to have found a more elegant and efficient solution. The solution, a version of which was first proposed in Chapman (1968), involves running in a way that seems to keep the ball moving at a constant speed through the visual field. As long as the fielder's own movements cancel any apparent changes in the ball's optical acceleration, she will end up in the location where the ball hits the ground. This solution, OAC (Optical Acceleration Cancellation), explains why fielders, when asked to stand still and simply predict where the ball will land, typically do rather badly. They are unable to predict the landing spot because OAC is a strategy that works by means of moment-by-moment self-corrections that, crucially, involve the agent's own movements. The suggestion that we rely on such a strategy is also confirmed by some interesting virtual reality experiments in which the ball's trajectory is suddenly altered in flight, in ways that could not happen in the real world—see Fink et al. 2009). OAC is a succinct case of fast, economical problem-solving. The canny use of data available in the optic flow enables the catcher to sidestep the need to deploy a rich inner model to calculate the forward trajectory of the ball.[12]

Such strategies are suggestive (see also Maturana & Varela 1980) of a very different role of the perceptual coupling itself. Instead of using sensing to get enough information inside, past the visual bottleneck, so as to allow the reasoning system to "throw away the world" and solve the problem wholly internally, such strategies use the sensor as *an open conduit allowing environmental magnitudes to exert a constant influence on behavior*. Sensing is here

12 There are related accounts of how dogs catch Frisbees—a rather more demanding task due to occasional dramatic fluctuations in the flight path (see Shaffer et al. 2004).

depicted as the opening of a channel, with successful whole-system behavior emerging when activity in this channel is kept within a certain range. In such cases:

> [T]he focus shifts from accurately representing an environment to continuously engaging that environment with a body so as to stabilize appropriate co-ordinated patterns of behaviour. (Beer 2000, p. 97)

These focal shifts may be fluidly accommodated within the PP framework. To see how, recall that "precision weighting" alters the gain on specific prediction error units, and thus provides a means of systematically varying the relative influence of different neural populations. The most familiar role of such manipulations is to vary the balance of influence between bottom-up sensory information and top-down model-based expectation. But another important role is the implementation of fluid and flexible forms of large-scale "gating" among neural populations. This works because very low-precision prediction errors will have little or no influence upon ongoing processing, and will fail to recruit or nuance higher-level representations. Altering the distribution of precision weightings thus amounts, as we saw above, to altering the "simplest circuit diagram" (Aertsen & Preißl 1991) for current processing. When combined with the complex, cascading forms of influence made available by the apparatus of top-down prediction, the result is an inner processing economy that is (see Clark in press) "maximally context-sensitive".

This suggests a new angle upon the outfielder's problem. Here too, already-active neural predictions and simple, rapidly-processed perceptual cues must work together (if PP is correct) to determine a pattern of precision-weightings for different prediction-error signals. This creates a pattern of effective connectivity (a temporary distributed circuit) and, within that circuit, it sets the balance between top-down and bottom-up modes of influence. In the case at hand, however, efficiency demands selecting a circuit in which visual sensing is used to cancel the optical acceleration of the fly ball. This means giving high weighting to the prediction errors associated with cancelling the vertical acceleration of the ball's optical projection, and (to put it bluntly) not caring very much about anything else. Apt precision weightings here function to select *what to predict* at any given moment. They may thus select a prelearnt, fast, low-cost strategy for solving a problem, as task and context dictate. Contextually-recruited patterns of precision weighting thus accomplish a form of set-selection or strategy switching—an effect already demonstrated in some simple simulations of cued reaching under the influence of changing tonic levels of dopamine firing—see Friston et al. (2012).

Fast, efficient solutions have also been proposed in the context of reasoning and choice. In an extensive literature concerning choice and decision-making, it has been common to distinguish between "model-based" and "model-free" approaches (see e.g., Dayan & Daw 2008; Dayan 2012; Wolpert et al. 2003). Model-based strategies rely, as their name suggests, on a model of the domain that includes information about how various states (worldly situations) are connected, thus allowing a kind of principled estimation (given some cost function) of the value of a putative action. Such approaches involve the acquisition and the (computationally challenging) deployment of fairly rich bodies of information concerning the structure of the task-domain. Model-free strategies, by contrast, are said to "learn action values directly, by trial and error, without building an explicit model of the environment, and thus retain no explicit estimate of the probabilities that govern state transitions" (Gläscher et al. 2010, p. 585). Such approaches implement "policies" that typically exploit simple cues and regularities while nonetheless delivering fluent, often rapid, response.

The model-based/model-free distinction is intuitive, and resonates with old (but increasingly discredited) dichotomies between reason and habit, and between analytic evaluation and emotion. But it seems likely that the image of parallel, functionally independent, neural subsystems will not stand the test of time. For example, a recent functional Magnetic Resonance Imaging (fMRI) study (Daw et al. 2011) sug-

gests that rather than thinking in terms of distinct (functionally isolated) model-based and model-free learning systems, we may need to posit a single "more integrated computational architecture" Daw et al. 2011, p. 1204), in which the different brain areas most commonly associated with model-based and model-free learning (pre-frontal cortex and dorsolateral striatum, respectively) *each* trade in both model-free and model-based modes of evaluations and do so "in proportions matching those that determine choice behavior" (Daw et al. 2011, p. 1209). Top-down information, Daw et al. (2011) suggest, might then control the way different strategies are combined in differing contexts for action and choice. Within the PP framework, this would follow from the embedding of shallow "model-free" responses within a deeper hierarchical generative model. By thus combining the two modes within an overarching model-based economy, inferential machinery can, by and large, identify the appropriate contexts in which to deploy the model-free ("habitual") schemes. "Model-based" and "model-free" modes of valuation and response, if this is correct, name extremes along a single continuum, and may appear in many mixtures and combinations determined by the task at hand.

This suggests a possible reworking of the popular suggestion (Kahneman 2011) that human reasoning involves the operation of two functionally distinct systems: one for fast, automatic, "habitual" response, and the other dedicated to slow, effortful, deliberative reasoning. Instead of a truly dichotomous inner organization, we may benefit from a richer form of organization in which fast, habitual, or heuristically-based modes of response are often the default, but within which a large variety of possible strategies may be available. Humans and other animals would thus deploy multiple—rich, frugal and all points in between—strategies defined across a fundamentally unified web of neural resources (for some preliminary exploration of this kind of more integrated space, see Pezzulo et al. 2013). Some of those strategies will involve the canny use of environmental structure – efficient embodied prediction machines, that is to say, will often deploy minimal neural models that benefit from repeated calls to world-altering action (as when we use a few taps of the smartphone to carry out a complex calculation).

Nor, finally, is there any fixed limit to the complexities of the possible strategic embeddings that might occur even within a single more integrated system. We might, for example, use some quick-and-dirty heuristic strategy to identify a context in which to use a richer one, or use intensive model-exploring strategies to identify a context in which a simpler one will do. From this emerging vantage point the very distinction between model-based and model-free response (and indeed between System 1 and System 2) looks increasingly shallow. These are now just convenient labels for different admixtures of resource and influence, each of which is recruited in the same general way as circumstances dictate.[13]

3.2 Being human

There is nothing specifically human, however, about the suite of mechanisms explored above. The basic elements of the predictive processing story, as Roepstorff (2013, p. 45) correctly notes, may be found in many types of organism and model-system. The neocortex (the layered structure housing cortical columns that provides the most compelling neural implementation for predictive processing machinery) displays some dramatic variations in size but is common to all mammals. What, then, makes us (superficially at least) so very different? What is it that allows us—unlike dogs, chimps, or dolphins—to latch on to distal hidden causes that include not just food, mates, and relative social rankings, but also neurons, predictive processing, Higgs bosons, and black holes?

One possibility (Conway & Christiansen 2001) is that adaptations of the human neural apparatus have somehow conspired to create, in us, an even more complex and context-flexible

13 Current thinking about switching between model-free and model-based strategies places them squarely in the context of hierarchical inference, through the use of "Bayesian parameter averaging". This essentially associates model-free schemes with simpler (less complex) lower levels of the hierarchy that may, at times, need to be contextualized by (more complex) higher levels.

hierarchical learning system than is found in other animals. Insofar as the predictive processing framework allows for rampant context-dependent influence within the distributed hierarchy, the same basic operating principles might (given a few new opportunities for routing and influence) result in the emergence of qualitatively novel forms of behavior and control. Such changes might explain why human agents display what Spivey (2007, p. 169) describes as an "exceptional sensitivity to hierarchical structure in *any* time-dependent signal".

Another (possibly linked, and certainly highly complementary) possibility involves a potent complex of features of human life, in particular our ability to engage in temporally co-coordinated social interaction (see Roepstorff et al. 2010) and our ability to construct artifacts and design environments. Some of these ingredients have emerged in other species too. But in the human case the whole mosaic comes together under the influence of flexible and structured symbolic language (this was the target of the Conway and Christiansen paper mentioned above) and an almost obsessive drive (Tomasello et al. 2005) to engage in shared cultural practices. We are thus able to redeploy our core cognitive skills in the transformative context of exposure to what Roepstorff et al. (2010) call "patterned sociocultural practices". These include the use of symbolic codes (encountered as "material symbols" (Clark 2006) and complex social routines (Hutchins 1995, 2014)—and more general, all the various ploys and strategies known as "cognitive niche construction" (see Clark 2008).

A simple example is the way that learning to perform mental arithmetic has been scaffolded, in some cultures, by the deliberate use of an abacus. Experience with patterns thus made available helps to install appreciation of many complex arithmetical operations and relations (for discussion of this, see Stigler 1984). The specific example does not matter very much, to be sure, but the general strategy does. In such cases, we structure (and repeatedly re-strutcture) our physical and social environments in ways that make available new knowledge and skills—see Landy & Goldstone (2005). Prediction-hungry brains, exposed in the course of embodied action to novel patterns of sensory stimulation, may thus acquire forms of knowledge that were genuinely out-of-reach prior to such physical-manipulation-based re-tuning of the generative model. Action and perception thus work together to reduce prediction error against the more slowly evolving backdrop of a culturally distributed process that spawns a succession of designed environments whose impact on the development (e.g., Smith & Gasser 2005) and unfolding (Hutchins 2014) of human thought and reason can hardly be overestimated.

To further appreciate the power and scope of such re-shaping, recall that the predictive brain is not doomed to deploy high-cost, model-rich strategies moment-by-moment in a demanding and time-pressured world. Instead, that very same apparatus supports the learning and contextually-determined deployment of low-cost strategies that make the most of body, world, and action. A maximally simple example is painting white lines along the edges of a winding cliff-top road. Such environmental alterations allow the driver to solve the complex problem of keeping the car on the road by (in part) predicting the ebb and flow of various simpler optical features and cues (see e.g., Land 2001). In such cases, we are building a better world in which to predict, while simultaneously structuring the world to cue the low-cost strategy at the right time.

3.3 Extending the predictive mind

All this suggests a very natural model of "extended cognition" (Clark & Chalmers 1998; Clark 2008), where this is simply the idea that bio-external structures and operations may sometimes form integral parts of an agent's cognitive routines. Nothing in the PP framework materially alters, as far as I can tell, the arguments previously presented, both pro and con, regarding the possibility and actuality of genuinely extended cognitive systems.[14] What PP

14 For a thorough rehearsal of the positive arguments, see Clark (2008). For critiques, see Rupert (2004, 2009), Adams & Aizawa (2001), and Adams & Aizawa (2008). For a rich sampling of the ongoing debate, see the essays in Menary (2010) and Estany & Sturm (2014).

does offer, however, is a specific and highly "extension-friendly" proposal concerning the shape of the specifically neural contribution to cognitive success. To see this, reflect on the fact that known external (e.g., environmental) operations provide—by partly constituting—additional strategies apt for the kind of "meta-model-based" selection described above. 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. Minimal internal models that involve calls to world-recruiting actions may thus be selected in the same way as a purely internal model. The availability of such strategies (of trading inner complexity against real-world action) is the hallmark of embodied prediction machines.

As a simple illustration, consider the work undertaken by Pezzulo et al. (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 such 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 (as discussed in previous sections) 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, an iPhone, 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. 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 "temporary task-specific device"—see Anderson et al. 2012) recruited as a function of context-varying estimations of uncertainty.

4 Conclusion: Towards a mature science of the embodied mind

By self-organizing around prediction error, and by learning a generative rather than a merely discriminative (i.e., pattern-classifying) model, these approaches realize many of the goals of previous work in artificial neural networks, robotics, dynamical systems theory, and classical cognitive science. They self-organize around prediction error signals, perform unsupervised learning using a multi-level architecture, and acquire a satisfying grip—courtesy of the problem decompositions enabled by their hierarchical form—upon structural relations within a domain. They do this, moreover, in ways that are firmly grounded in the patterns of sensorimotor experience that structure learning, using continuous, non-linguaform, inner encodings (probability density functions and probabilistic inference). Precision-based restructuring of patterns of effective connectivity then allow us to nest simplicity within complexity, and to make as much (or as little) use of body and world as task and context dictate.

This is encouraging. It might even be that models in this broad ballpark offer us a first glimpse of the shape of a fundamental and unified science of the embodied mind.

Acknowledgements

This work was supported in part by the AHRC-funded 'Extended Knowledge' project, based at the Eidyn research centre, University of Edinburgh.

References

Adams, F. & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14* (1), 43-64. 10.1080/09515080120033571

——— (2008). *The bounds of cognition.* Malden, MA: Blackwell Publishing.

Adams, R. A., Perrinet, L. U. & Friston, K. (2012). Smooth pursuit and visual occlusion: Active inference and oculomotor control in schizophrenia. *PLoS One*, *7* (10), e47502. 10.1371/journal.pone. 0047502

Adams, R. A., Shipp, S. & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218* (3), 611-643. 10.1007/s00429-012-0475-5

Aertsen, A. & Preißl, H. (1991). Dynamics of activity and connectivity in physiological neuronal networks. In H. G. Schuster (Ed.) *Nonlinear dynamics and neuronal networks* (pp. 281-302). Weinheim, GER: VCH Verlag.

Anderson, M. L., Richardson, M. & Chemero, A. (2012). Eroding the boundaries of cognition: Implications of embodiment. *Topics in Cognitive Science*, *4* (4), 717-730. 10.1111/j.1756-8765.2012.01211.x

Anscombe, G. E. M. (1957). *Intention.* Oxford, UK: Basil Blackwell.

Ballard, D. (1991). Animate vision. *Artificial Intelligence*, *48*, 57-86. 10.1016/0004-3702(91)90080-4

Ballard, D., Hayhoe, M., Pook, P. & Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20* (4), 723-767.

Bastian, A. (2006). Learning to predict the future: The cerebellum adapts feedforward movement control. *Current opinion in neurobiology*, *16* (6), 645-649.

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* (4), 695-711. 10.1016/j.neuron.2012.10.038

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., De Weerd, P., Kennedy, H. & Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron*. 10.1016/j.neuron.2014.12.018

Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, *4* (3), 91-99. 10.1016/S1364-6613(99)01440-0

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139-159. 10.1.1.12.1680

Brown, H., Adams, R. A., Parees, I., Edwards, M. & Friston, K. (2013). Active inference, sensory attenuation

and illusions. *Cognitive Processing*, *14* (4), 411-427. 10.1007/s10339-013-0571-3

Chapman, S. (1968). Catching a baseball. *American Journal of Physics*, *36*, 868-870.

Churchland, P. S., Ramachandran, V. S. & Sejnowski, T. J. (1994). A critique of pure vision. In C. Koch & J. L. Davis (Eds.) *Large Scale Neuronal Theories of the Brain* (pp. 23-60). Cambridge, MA: MIT Press.

Clark, A. (2006). Language, embodiment and the cognitive niche. *Trends in Cognitive Sciences*, *10* (8), 370-374. 10.1016/j.tics.2006.06.012

——— (2008). *Supersizing the mind: Action, embodiment, and cognitive extension.* New York, NY: Oxford University Press.

——— (2013a). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36* (3), 181-204. 10.1017/S0140525X12000477

——— (2013b). The many faces of precision. *Frontiers in Theoretical and Philosophical Psychology*, *4* (270), 1-9. 10.3389/fpsyg.2013.00270

——— (in press). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind.* New York, NY: Oxford University Press.

Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis*, *58* (1), 7-19. 10.1111/1467-8284.00096

Conway, C. & Christiansen, M. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, *5* (12), 539-546. 10.1016/S1364-6613(00)01800-3

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*, 1204-1215. 10.1016/j.neuron.2011.02.02

Dayan, P. (2012). How to set the switches on this thing. *Current Opinion in Neurobiology*, *22* (6), 1068-1074. 10.1016/j.conb.2012.05.011

Dayan, P. & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, *8* (4), 429-453. 10.3758/CABN.8.4.429

Edwards, M. J., Adams, R. A., Brown, H., Pareés, I. & Friston, K. (2012). A Bayesian account of 'hysteria'. *Brain*, *135* (11), 3495-3512. 10.1093/brain/aws129

Egner, T., Monti, J. M. & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience*, *30* (49), 16601-16608. 10.1523/JNEUROSCI.2770-10.2010

Estany, A. & Sturm, T. (Eds.) (2014). Extended cognition: New philosophical perspectives. *Special Issue of Philosophical Psychology*, *27* (1)

Feldman, A. G. (2009). New insights into action-perception coupling. *Experimental Brain Research*, *194* (1), 39-58. 10.1007/s00221-008-1667-3

Feldman, H. & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4* (215), 1-23. 10.3389/fnhum.2010.00215

Felleman, D. J. & Van Essen, D. C. (1991). Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex*, *1* (1), 1-47. 10.1093/cercor/1.1.1-a

Fink, P. W., Foo, P. S. & Warren, W. H. (2009). Catching fly balls in virtual reality: A critical test of the outfielder problem. *Journal of Vision*, *9* (13), 1-8. 10.1167/9.13.14

Flash, T. & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *The Journal of Neuroscience*, *5* (7), 1688-1703. 10.1.1.134.529

Fletcher, P. & Frith, C. (2009). Perceiving is believing: A Bayesian appraoch to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*, 48-58. 10.1038/nrn2536

Franklin, D. W. & Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, *72* (3), 425-442. 10.1016/j.neuron.2011.10.006

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences 29*, *360* (1456), 815-836. 10.1098/rstb.2005.1622

——— (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4* (11), e1000211. 10.1371/journal.pcbi.1000211

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

——— (2011). What is optimal about motor control? *Neuron*, *72* (3), 488-498. 10.1016/j.neuron.2011.10.018

Friston, K. J., Daunizeau, J., Kilner, J. & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, *102* (3), 227-260. 10.1007/s00422-010-0364-z

Friston, K., Mattout, J. & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, *104* (1-2), 137-160. 10.1007/s00422-011-0424-z

Friston, K., Adams, R. A., Perrinet, L. & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, *3* (151), 1-20. 10.3389/fpsyg.2012.00151

Friston, K., Samothrakis, S. & Montague, R. (2012). Active inference and agency: Optimal control without cost functions. *Biological Cybernetics*, *106* (8-9), 523-541. 10.1007/s00422-012-0512-8

Friston, K. J., Shiner, T., Fitzgerald, T., Galea, J. M., Adams, R., Brown, H., Dolan, R. J., Moran, R., Stephan, K. E. & Bestmann, S. (2012). Dopamine, affordance and active inference. *PLoS Computational Biology*, *8* (1), e1002327. 10.1371/journal.pcbi.1002327

Frith, C. D. & Friston, K. J. (2012). False perceptions and false beliefs: Understanding schizophrenia. *Working Group on Neurosciences and the Human Person: New Perspectives on Human Activities, The Pontifical academy of Sciences, 8-10 November 2012.* Vatican City, VA: Casina Pio IV.

Gläscher, J., Daw, N., Dayan, P. & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model based and model-free reinforcement learning. *Neuron*, *66* (4), 585-595. 10.1016/j.neuron.2010.04.016

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27* (3), 377-442. 10.1017/S0140525X04000093

Haruno, M., Wolpert, D. M. & Kawato, M. (2003). Hierarchical MOSAIC for movement generation. *International congress series*, *1250*, 575-590.

Hinton, G. E., Osindero, S. & Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18* (7), 1527-1554. 10.1162/neco.2006.18.7.1527

Hinton, G. E. & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313* (5786), 504-507. 10.1126/science.1127647

Hohwy, J. (2013). *The predictive mind.* New York, NY: Oxford University Press.

——— (2014). The neural organ explains the mind. In T. Metzinger & J. M. Windt (Eds.) *Open MIND.* Frankfurt a. M., GER: MIND Group.

Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: MIT Press.

——— (2014). The cultural ecosystem of human cognition. *Philosophical Psychology*, *27* (1), 34-49. 10.1080/09515089.2013.830548

James, W. (1890). *The principles of psychology Vol. I, II.* Cambridge, MA: Harvard University Press.

Jordan, M. I. & Jacobs, R. A. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, *6* (2), 181-214. 10.1162/neco.1994.6.2.181

Kahneman, D. (2011). *Thinking fast and slow.* London, UK: Penguin.

Kawato, K. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *9* (6), 718-727. 10.1016/S0959-4388(99)00028-8

Land, M. (2001). Does steering a car involve perception of the velocity flow field? In J. M. Zanker & J. Zeil (Eds.) *Motion vision - Computational, neural, and ecological constraints* (pp. 227-238). Berlin, GER: Springer Verlag.

Landy, D. & Goldstone, R. L. (2005). How we learn about things we don't already understand. *Journal of Experimental and Theoretical Artificial Intelligence*, *17* (4), 343-369. 10.1080/09528130500283832

Littman, M., Majercik, S. & Pitassi, T. (2001). Stochastic Boolean satisfiability. *Journal of Automated Reasoning*, *27* (3), 251-296.

Lotze, H. (1852). *Medizinische Psychologie oder Physiologie der Seele.* Leipzig, GER: Weidmannsche Buchhandlung.

Marr, D. (1982). *Vision: A computational approach.* San-Francisco, CA: Freeman & Co.

Maturana, H. & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living.* Boston, MA: Reidel.

McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks: A historical and tutorial review. *Frontiers in Psychology*, *4* (503), 1-25. 10.3389/fpsyg.2013.00503

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

Mohan, V., Morasso, P., Metta, G. & Kasderidis, S. (2010). Actions & imagined actions in cognitive robots. In V. Cutsuridis, A. Hussain & J. G. Taylor (Eds.) *Perception-reason-action cycle: Models, architectures, and hardware* (pp. 1-32). New York, NY: Springer Series in Cognitive and Neural Systems.

Mohan, V. & Morasso, P. (2011). Passive motion paradigm: An alternative to optimal control. *Frontiers in Neurorobotics*, *5* (4), 1-28. 10.3389/fnbot.2011.00004

Namikawa, J., Nishimoto, R. & Tani, J. (2011). A neurodynamic account of spontaneous behaviour. *PLoS Computational Biology*, *7* (10), e100222. 10.1371/journal.pcbi.1002221

Namikawa, J. & Tani, J. (2010). Learning to imitate stochastic time series in a compositional way by chaos. *Neural Networks*, *23* (5), 625-638. 10.1016/j.neunet.2009.12.006

Park, J. C., Lim, J. H., Choi, H. & Kim, D. S. (2012). Predictive coding strategies for developmental neurorobotics. *Frontiers in Psychology*, *3* (134), 1-10. 10.3389/fpsyg.2012.00134

Pellicano, E. & Burr, D. (2012). When the world becomes too real: A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16* (10), 504-510. 10.1016/j.tics.2012.08.009

Pezzulo, G. (2008). Coordinating with the future: The anticipatory nature of representation. *Minds and Machines*, *18*, 179-225. 10.1007/s11023-008-9095-5

——— (2014). Why do you fear the bogeyman? An embodied predictive coding model of perceptual inference. *Cognitive, Affective, & Behavioral Neuroscience*, *14* (3), 902-911. 10.3758/s13415-013-0227-x

Pezzulo, G., Barsalou, L., Cangelosi, A., Fischer, M., McRae, K. & Spivey, M. (2013). Computational grounded cognition: A new alliance between grounded cognition and computational modeling. *Frontiers in Psychology*, *3* (612), 1-11. 10.3389/fpsyg.2012.00612

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* (92), 1-15. 10.3389/fpsyg.2013.00092

Poeppel, D. & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes*, *26* (7), 935-951. 10.1080/01690965.2010.493301

Price, C. J. & Devlin, J. T. (2011). The interactive Account of ventral occipito-temporal contributions to reading. *Trends in Cognitive Sciences*, *15* (6), 246-253. 10.1016/j.tics.2011.04.001

Raichle, M. E. & Snyder, A. Z. (2007). A default mode of brain function: A brief history of an evolving idea. *NeuroImage*, *37* (4), 1083-1090. 10.1016/j.neuroimage.2007.02.041

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), 79-87. 10.1038/4580

Roepstorff, A. (2013). Interactively human: Sharing time, constructing materiality: Commentary on Clark. *Behavioral and Brain Sciences*, *36* (3), 224-225. 10.1017/S0140525X12002427

Roepstorff, A., Niewöhner, J. & Beck, S. (2010). Enculturating brains through patterned practices. *Neural Networks*, *23*, 1051-1059. 10.1016/j.neunet.2010.08.002

Roth, M. J., Synofzik, M. & Lindner, A. (2013). The cerebellum optimizes perceptual predictions about external sensory events. *Current Biology*, *23* (10), 930-935. 10.1016/j.cub.2013.04.027

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

——— (2009). *Cognitive systems and the extended mind.* Oxford, UK: Oxford University Press.

Saegusa, R., Sakka, S., Metta, G. & Sandini, G. (2008). *Sensory prediction learning - how to model the self and environment.* Annecy, FR: The 12th IMEKO TC1-TC7 joint Symposium on "Man Science and Measurement" (IMEKO2008).

Seth, A. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17* (11), 565-573. 10.1016/j.tics.2013.09.007

Shaffer, D. M., Krauchunas, S. M., Eddy, M. & McBeath, M. K. (2004). How dogs navigate to catch frisbees. *Psychological Science*, *15* (7), 437-441. 10.1111/j.0956-7976.2004.00698.x

Shea, N. (2013). Perception vs. action: The computations may be the same but the direction of fit differs: Commentary on Clark. *Behavioral and Brain Sciences*, *36* (3), 228-229. 10.1017/S0140525X12002397

Shipp, S., Adams, R. A. & Friston, K. J. (2013). Reflections on agranular architecture: Predictive coding in the motor cortex. *Trends in Neurosciences*, *36* (12), 706-716. 10.1016/j.tins.2013.09.004

Smith, L. & Gasser, M. (2005). The development of embodied cognition: Six lessons from babies. *Artificial Life*, *11*, 13-29. 10.1162/1064546053278973

Sommer, M. A. & Wurtz, R. H. (2006). Influence of thalamus on spatial visual processing in frontal cortex. *Nature*, *444* (7117), 374-377. 10.1038/nature05279

——— (2008). Brain circuits for the internal monitoring of movements. *Annual Review of Neuroscience*, *31* (1), 317-338. 10.1146/annurev.neuro.31.060407.125627

Spivey, M. J. (2007). *The continuity of mind.* New York, NY: Oxford University Press.

Spratling, M. W. (2008). Predictive coding as a model of biased competition in visual attention. *Annual Review of Neuroscience*, *48* (12), 1391-1408. 10.1146/annurev.neuro.31.060407.125627

——— (2010). Predictive coding as a model of response properties in cortical area V1. *The Journal of Neuroscience*, *30* (9), 3531-3543. 10.1523/JNEUROSCI.4911-09.2010

——— (2013). Distinguishing theory from implementation in predictive coding accounts of brain function. *Behavioral and Brain Sciences*, *36* (3), 231-232.

Stigler, J. W. (1984). "Mental abacus": The effect of abacus training on Chinese children mental calculation. *Cognitive Psychology*, *16* (2), 145-176. 10.1016/0010-0285(84)90006-9

Tani, J. (2007). On the interactions between top-down anticipation and bottom-up regression. *Frontiers in Neurorobotics*, *1* (2), 2. 10.3389/neuro.12.002.2007

Thelen, E. & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action.* Massachusetts, MA: MIT Press.

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, *7* (9), 907-915. 10.1038/nn1309

Todorov, E. & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, *5* (11), 1226-1235. 10.1038/nn963

Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005). Understanding and sharing intentions: The ontogeny and phylogeny of cultural cognition. *Behavioral and Brain Sciences*, *28* (5), 675-691. 10.1017/S0140525X05000129

Uno, Y., Kawato, M. & Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement. *Biological Cybernetics*, *61* (2), 89-101. 10.1007/BF00204593

Von Holst, E. (1954). "Relations between the central Nervous System and the peripheral organs". *The British Journal of Animal Behaviour*, *2* (3), 89-94. 10.1016/S0950-5601(54)80044-X

Wolpert, D. M., Doya, K. & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London*, *358* (1431), 593-602. 10.1098/rstb.2002.1238

Wolpert, D. M. & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11* (7-8), 1317-1329. 10.1016/S0893-6080(98)00066-5

Wolpert, M. & Miall, D. M. (1996). Forward Models for Physiological Motor Control. *Neural Networks*, *9* (8), 1265-1279.

Wurtz, R. H., McAlonan, K., Cavanaugh, J. & Berman, R. A. (2011). Thalamic pathways for active vision. *Trends in Cognitive Sciences*, *15* (4), 177-184. 10.1016/j.tics.2011.02.004

Yamashita, Y. & Tani, J. (2008). Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment. *PLoS ONE*, *6* (10), e1000220. 10.1371/annotation/c580e39c-00bc-43a2-9b15-af71350f9d43

Zorzi, M., Testolin, A. & Stoianov, I. (2013). Modeling language and cognition with deep unsupervised learning: a tutorial overview. *Frontiers Psychology*, *4* (415), 1-14. 10.3389/fpsyg.2013.00515

Extending the Explanandum for Predictive Processing

A Commentary on Andy Clark

Michael Madary

In this commentary, I suggest that the predictive processing framework (PP) might be applicable to areas beyond those identified by Clark. In particular, PP may be relevant for our understanding of perceptual content, consciousness, and for applied cognitive neuroscience. My main claim for each area is as follows:

1) PP urges an organism-relative conception of perceptual content.
2) Historical *a priori* accounts of the structure of perceptual experience converge with results from PP.
3) There are a number of areas in which PP can find important practical applications, including education, public policy, and social interaction.

Keywords
Anticipation | Applied cognitive neuroscience | Consciousness | Perception | Perceptual content | Phenomenology | Predictive processing

Commentator

Michael Madary
madary@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Target Author

Andy Clark
Andy.Clark@ed.ac.uk
University of Edinburgh
Edinburgh, United Kingdom

Editors

Thomas Metzinger
metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt
jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 Introduction

An understandable reaction to the predictive processing framework (PP) is to think that it is too ambitious (Hohwy this collection). My suggestion in this commentary is the opposite. I will argue that PP can be fruitfully applied to areas of inquiry that have so far received little, if any, attention from the proponents of PP. Perhaps we can extend the explanandum even further than Andy Clark has recommended.

There is a certain rhetorical danger to the position I am urging. One should not oversell one's case. I hope to avoid this danger by being clear upfront that my goal is not to convince the skeptic of the attraction of PP. I cannot improve on Clark (and others, see below) in that regard. Instead, I investigate the following question: if some version of PP (again, see below) is true, then what are the larger implications for human self-understanding? My answer to this question covers three topics. First I will engage with Clark's discussion of perceptual processing from sections 1 and 2.1 of his article. There I

will sketch how PP's reversal of the traditional model of perceptual processing may have significant implications for the way in which we understand perceptual content, which is a core issue in the philosophy of psychology. In the second section I will turn to another area of philosophical concern: consciousness. Historically, consciousness research has had a rocky relationship with the sciences of the mind. I hope to point towards the possibility of a rapprochement. In the final section of the commentary, I will quickly touch on some practical matters. If PP is true, then there are important consequences for the way in which we approach topics in education, public policy, and social interaction.

My goal is to indicate possible areas in which Clark's article (and related themes) might serve as a foundation for future directions of research. My main claims are as follows, numbered according to each section:

1. PP urges an organism-relative conception of perceptual content.
2. Historical *a priori* accounts of the structure of perceptual experience converge with results from PP.
3. There are a number of areas in which PP can find important practical applications.

Before entering into the specific issues, I should add a note about what I mean by PP. Here I am following the general theoretical framework expressed in Clark's article as well as in a number of other publications (Clark 2013; Hohwy 2013). The approach has a number of intellectual roots, including Hermann von Helmholtz (1867) and Richard Gregory (1980). The main contemporary expression of PP perhaps owes the most to Karl Friston (2005, 2008, 2010) and his collaborators, also with important developments of the generative model by Geoffrey Hinton (2007). By referring to PP as one general framework, I do not mean to imply that there are no outstanding issues of disagreement or open questions within PP. As Clark indicates, citing Spratling (2013), there are a number of options being developed as to the specific implementation of PP. Also, in the philosophical literature there is an emerging question about whether to understand PP as internalist or externalist regarding the vehicles of mental states (Hohwy 2014)—I take no position either way here, but see footnote 2. Overall, my remarks are motivated by Clark's exposition of PP, but they should be applicable to other approaches and interpretations as well.

2 A new conception of perceptual content

Clark has emphasized the way in which PP departs from the standard picture in perceptual psychology, and from David Marr's (1982) model of visual processing in particular (pp. 1–5). According to the standard account, the flow of information is "bottom-up," as perceptual systems construct increasingly sophisticated representations based on the information transduced at the periphery. According to PP, perception involves the active prediction of the upcoming sensory input, "top-down." Deviation from what is predicted, known as the prediction error, propagates upwards through the hierarchy until it is explained away by the Bayesian generative model.

Now I would like to add that the standard picture in perceptual psychology has been widely regarded as complementary to the standard picture in the philosophy of perception (see Tye 2000, for example). One central question in the philosophy of perception is the following: what is the *content* of perceptual states? Or, what does perception *represent*? The standard answer, in tune with Marr's approach, is that perceptual systems represent the external world, more or less as it really is. As Marr puts it, the purpose of vision is "to know what is where by looking" (1982). This way of thinking about perceptual content is almost a commonplace in the philosophical literature (Lewis 1980, p. 239; Fodor 1987, Ch. 4; Dretske 1995, Ch. 1). Kathleen Akins has described how the orthodox conception regards the senses as "servile" in that they report on the environmental stimulus "without fiction or embellishment" (1996, pp. 350–351).

Since PP overturns the reigning model in perceptual psychology, one might now ask

whether it also overturns the reigning model in the philosophy of perception. Here are two initial reasons to think that it does. First, according to PP, there is always an active contribution from the organism, or at least from a part of the organism. Perceptual states are generated internally and spontaneously by the ongoing dynamics of the generative model. Those states are *constrained* by perceptual sampling of the world, not driven by input from the world. Perceptual states are driven by the endogenous activity of the predictive brain. The relevant causal history of these states begins, if you will, within the brain, rather than from the outside. Each organism's generative model is unique in that it has been formed and continuously revised according to the particular trajectory of that organism's cycle of action and perception. As Clark himself puts it, the forward flow of sensory information is always "*relative to specific predictions*" (p. 6). These considerations make it clear that there can be variation in perceptual content for identical environmental conditions. Perceivers with different histories will have different predictions (Madary 2013, pp. 342–345). The degree of variation is an open question, but it is reasonable to expect variation.

A second reason to think that PP motivates a richer conception of perceptual content is that perception, according to PP, is not simply in the service of informing the organism "what is where." One main feature of PP is that perception and action work together in the service of minimizing prediction error. Clark explains that in "active inference [...] the agent moves its sensors in ways that amount to actively seeking or generating the sensory consequences that they [...] expect" (2013, p. 6, also see his discussion on page 16). If this is right, then perception does not serve the purpose of simply reporting on the state of the environment. Instead, perception is guided by expectation. While the received view of perceptual content answers the question of "what is out there?", PP suggests that perceptual content answers the question of "is this what I expected and tested via active inference?" In a way, PP simplifies perceptual content by replacing the goal of representing the world with the single guiding principle of error minimization.

These two points suggest an understanding of perceptual content as something that is deeply informed by the specific history and embodiment of the organism. The content of perception is a complex interplay between particular organisms and their particular environments. At least on the face of it, this way of considering perception suggests new challenges and interesting new theoretical options for philosophers interested in describing perceptual content. For one thing, it suggests that propositional content as expressed using natural language (Searle 1983, p. 40) may be ill-suited for the task of describing perceptual content. Natural language does not typically include reports about prediction-error minimization, nor does it capture the fine-grained differences in perceptual content that will arise due to slight variations in the predictions made by different organisms. The traditional account of perceptual content, following Marr, does not include such differences, and is thus better disposed to expression using natural language.

These new challenges for understanding perceptual content may offer at the same time a general lesson for understanding all mental content in a naturalistic manner. Let me explain. One of the main goals in the philosophy of psychology has been to naturalize intentionality, to give an account of the content of mental states in terms of the natural sciences (in non-mentalistic terms). Well-known attempts include causal co-variation (Fodor 1987, Ch. 4) and teleosemantics (Millikan 1984, 2004). All attempts have met with compelling counterexamples.[1] Importantly, one implicit presupposition in the debate is that mental content should be conceived along the lines of the traditional view of perceptual content sketched above. That is, mental states are thought to be about bits of the objective world considered independently of the particular organism who possesses those mental states. To use a standard example, my belief that there is milk in the refrigerator is true if and only if there is milk in the refriger-

1 For an overview of the major theories and their challenges, see Jacob (2010, section 9) and the references therein.

ator. This belief is about bits of the objective world: milk and the refrigerator in particular. Nothing else about my mind is deemed relevant for understanding the content of that belief. To use the familiar phrase, beliefs have a mind-to-world direction of fit (based on Anscombe 1957, §32).

If my reading of PP is right, and perceptual content turns out to be a matter of the complex interaction between particular organisms and their environments, then the comfortable pre-theoretical mind/world distinction might need revision.[2] Recall the discussion above, in which I claimed that, on the new PP-inspired understanding of perception the question is about whether sensory stimulation fulfils the expectations of particular organisms. All perceptual states are thereby colored, as it were, by the mental lives of the organisms having those states. Organisms are not interested in what the world is like. Organisms are interested in sustaining their integrity and physical existence; they are interested in what the world is like *relative to their own particular sensorimotor trajectory through the world*, a trajectory that is partly determined by their phenotype (Friston et al. 2006). This refashioning of the mind/world relationship is unorthodox, but it is hardly new. Similar ideas can be found in von Uexküll's *Umwelt* (1934), Merleau-Ponty's discussion of sensory stimuli (1962, p. 79), Millikan's "pushmi-pullyu" representations (1995), Akins' narcissistic sensory systems (1996), Clark's earlier work (1997, Ch. 1), and in Metzinger's ego tunnel (2009, pp. 8–9).

Now return to the problem of naturalizing intentionality. If we replace the notion of a purely world-directed mental state with a world-relative-to-the-organism-directed mental state, then naturalizing intentionality must somehow incorporate the relationship between the organism and its world. One way to pursue this project is to make it a matter of biology and physics. All living organisms keep themselves far from thermodynamic equilibrium by continuously exchanging matter and energy with their environment (Haynie 2008). Perhaps intentionality can be recast in terms of the organism's ongoing struggle to maintain itself as a living entity. This line of thought is central to the enactivist "sense-making" of Maturana, Varela, and Thompson (Maturana & Varela 1980; Thompson 2007). Crucially, it is also a central feature of Friston's version of PP. According to Friston, prediction error minimization is a kind of functional description for the physical process of the organism's minimizing free energy in its effort to maintain itself far from thermodynamic equilibrium (2013). Naturalizing intentionality may be just a matter of physics (see Dixon et al. 2014 for an implementation of this strategy for problem-solving tasks).

Before moving on to the next section, I should add two qualifications. First, the idea of perceptual content being partly determined by the particular history of the perceiver should not be misunderstood as some kind of radical relativism with regard to perceptual content. Even if perceptual content is *partly* determined by the details of the organism, it is also partly determined by the world itself. As proponents of PP frequently claim, our generative models mirror the causal structure of the world (Hohwy 2013, Ch. 1). The point I am emphasizing here is that the causal structure of the world that is extracted is a structure relative to the embodiment (see Clark this collection, section 2.4)—and perceptual history—of the perceiver. The causal structure mirrored by a chimpanzee's generative model is, in important ways, unlike the causal structure mirrored by that of a catfish.

The second qualification has to do with my remark that naturalizing intentionality may be just a matter of physics. Even if one allows that the approach I sketched shows promise, it is important to emphasize the explanatory gulf that remains. The intentionality-as-physics approach might succeed in explaining a bacterium's intentional directedness towards a

2 One possibility here has been explored recently by Karl Friston using the concept of a Markov blanket, which produces a kind of partition between information states. As I read Friston, he advocates a pluralism about Markov blankets. On this view, there is not one boundary between mind and world, but instead there are a number of salient boundaries within, and perhaps around, living organisms. Friston writes that " . . . a system can have a multitude of partitions and Markov blankets . . . the Markov blanket of an animal encloses the Markov blankets of its organs, which enclose Markov blankets of cells, which enclose Markov blankets of nuclei . . ." (2013, p. 10).

sugar gradient (Thompson 2007, p. 74–75), but it is far from clear how it would apply to my belief that P—say, for example, that California Chrome won the Kentucky Derby in 2014.

The main argument of this section has been that PP motivates an understanding of perceptual content that is always organism-relative. Clark's version of PP, while not in conflict with this idea, has not addressed it explicitly, especially as it relates to the philosophy of perception. My goal here has been to do just that.

3 Consciousness

In this section I would like to consider how conscious experience might relate to the PP framework. In particular, I suggest that there is a convergence between *a priori* descriptions of consciousness, on one hand, and the structure of information processing according to PP on the other.[3] I will not remark on the way in which PP relates to some well-known issues in the study of consciousness, such as the hard problem or the explanatory gap. It is not clear to me that PP has anything new to contribute to these topics. Nor will I make any claims about which existing theories of the neural basis of consciousness fit best with PP, although I suspect there is some interesting work there to be done.

My main concern here is in the *structure* of conscious experience, of visual experience in particular. Here I adopt a strategy recommended by Thomas Nagel (1974), and David Chalmers (1996, pp. 224–225). Nagel puts the idea nicely, "[…] structural features of perception might be more accessible to objective description, even though something would be left out" (1974, p. 449, cited in Chalmers 1996, pp. 382 f.). The strategy has been implemented, in fact, using Marr's theory of vision—the theory that, as Clark puts it, PP turns upside down. Ray Jackendoff (1987, p. 178) and Jesse Prinz (2012, p. 52) have both emphasized the structural similarities between conscious visual experience and Marr's 2.5 dimensional sketch. Visual phenomenology is not a flat two-dimensional surface, because we see depth. But neither is visual phenomenology fully three-dimensional, because we cannot see the hidden sides of objects. Marr's 2.5 dimensional representation captures the level in-between two and three dimensional representation that seems to correspond to our visual phenomenology; it captures Hume's insight that visual experience is perspectival: "The table, which we see, seems to diminish, as we remove farther from it […]" (1993, p. 104).

As Hume emphasized the perspectival nature of visual experience, Kant famously emphasized the temporal nature of experience in the second section of the Transcendental Aesthetic: "Time is a necessary representation (*Vorstellung*), which lays at the foundation of all intuitions" (1781/1887/1998, A31). In an elegant synthesis of these two features of visual experience, Edmund Husserl suggested that the general structure of visual experience is one of anticipation and fulfillment:

> Every percept, and every perceptual context, reveals itself, on closer analysis, as made up of components which are to be understood as ranged under two standpoints of intention and (actual or possible) fulfillment. (Logical Investigation, VI §10 1900, Findlay trans., 1970)

In this passage from his early work, Husserl writes of "intention and fulfillment," but he later replaced "intention" with "anticipation" when dealing with perception.[4]

The main point is fairly straightforward: we perceive properties by implicitly anticipating how the appearances of those properties will

3 For a theoretical treatment of the functional significance of this convergence, see Metzinger & Gallese (2003).

4 When first developing the framework, he used the more general term "intention" because he was dealing with linguistic meaning, not perception. When applying the framework to perception one can be more precise about the nature of the empty perceptual intentions: they are anticipatory. In his later work, his *Analyses of Passive Synthesis* from the 1920s, Husserl ties in perceptual intentions with his work on time consciousness (1969) and refers to them as protentions (*Protentionen*; Husserl 1966, p. 7). In the same work, he refers to perceptual protentions as anticipations (*Erwartungen,* 1966, p. 13, and *antizipiert*, 1966, p. 7). See Madary (2012a) for a discussion of how Husserl's framework can be situated relative to contemporary philosophy of perception. Also see Bernet et al. (1993, p. 128) and Hopp (2011).

change as we move (or as the objects move). Husserl's proposal accommodates the perspectival character of experience because it addresses the question of how we perceive objective properties despite being constrained to one perspective at a time. And it accommodates the temporal nature of experience because anticipation is always future-directed.

Here is not the place to enter into the details of the thesis that the general structure of conscious experience is one of anticipation and fulfillment (see my 2013 for some of these details), but I should add one more point. As both Husserl (1973, p. 294) and Daniel Dennett (1991, Ch. 3) have noted, peripheral vision is highly indeterminate.[5] Also, as we explore our environment we experience a continuous trade off between determinacy and indeterminacy. As I lean in for a closer look at one object, the other objects in my visual field fade into indeterminacy. In order to account for this feature of experience, we can note that visual anticipations have various degrees of determinacy.[6]

Now let us return to PP. *If Hume provides the philosophy of perception for Marr's theory of vision, then Husserl provides the philosophy of perception for PP.* The structural similarities should be apparent. The predictive brain underlies the essentially anticipatory structure of perceptual awareness. Degrees of determinacy are encoded probabilistically in our generative models (Clark 2013; Madary 2012b). Action and perception are tightly linked (Clark this collection, p. 9) as self-generated movements stir up new perceptual anticipations.

Many readers will see a connection between the thesis of anticipation and fulfillment, on one hand, and the sensorimotor approach to perception (O'Regan & Noë 2001; Noë 2004) on the other. Overall, there is significant thematic overlap between the two (Madary 2012a, p. 149). As Seth (2014) has argued, many of the central claims of the sensorimotor approach can be incorporated into the PP framework.[7] This synthesis offers impressive explanatory power, bringing the standard sensorimotor experimental evidence (reversing goggles, change blindness, selective rearing) together with the theoretical neuroscience of PP. The explanatory power is even more impressive if I am correct that PP reflects the general structure of visual phenomenology, where predictive processing corresponds to perceptual anticipations and probabilistic coding corresponds to experienced indeterminacy.

4 Applied cognitive neuroscience

I would like to begin this section with some general comments about new opportunities for human self-understanding, about extending the explanandum. Academic disciplines are standardly divided into the sciences and the humanities, and some have expressed discomfort about the distance between the two modes of inquiry, or between the two cultures, as Snow (1959) famously put it (also see Brockman 1996). There is an immediate appeal to Metzinger's assertion that "Epistemic progress in the real world is something that is achieved by all disciplines *together*" (2003, p. 4). *If my claims from the previous section are on the right track, then we have a convergence of results between the two independent modes of inquiry, between the empirical sciences and the humanities.* It is tempting to hope that this convergence signals the beginning of a rapprochement between the sciences and the humanities. Perhaps we are at the threshold of a new science of the mind (Rowlands 2010), a science that finds natural and fruitful connections with the world of human experience. In this section, I will explore possible connections with education, public policy, and social interaction.

Clark makes two main claims in the final sections of his article that serve for the basis of my comments here. First, he suggests that PP motivates an understanding of cognitive processing as "maximally context sensitive" (p. 16), which follows from the property of PP systems being highly flexible in setting precision weightings for the incoming prediction errors. Flexibility in weighting precision enables flexibility in the deployment of processing resources. Thus there may be a wide variety of cognitive strategies at our disposal, with a continuous in-

5 For impressive empirical work on this theme, see Freeman & Simoncelli (2011).

terplay between more costly and less costly strategies. Second, he addresses the challenge of explaining why humans have unique cognitive powers unavailable to non-human animals who have the same fundamental PP architecture. In response to this challenge, Clark suggests that our abilities may be due to our patterns of social interaction as well as our construction of artifacts and "designer environments" (p. 19). Taken together, these two claims can be used to inform practical decisions in a number of ways.

Begin with education. Educational psychology is a broad and important area of research. PP suggests new ways of approaching human learning, ways that might depart from the received views that have guided educational psychology. I cannot begin to engage with this huge issue here, but I would like to offer one quick example. One fairly well-known application of educational psychology is in the concept of scaffolded learning, which is built on work by Lev Vygotsky and Jerome Bruner. As it is used now, scaffolded learning involves providing the student with helpful aids at particular stages of the learning process. These aids could include having a teacher present to give helpful hints, working in small groups, and various artifacts designed with the intention of anticipating stages at which the student will need help, such as visual aids, models, or tools. Clark himself mentions the abacus, which is central example of scaffolded learning (p. 19). More generally, scaffolded learning is a good example of what Richard Menary has called "cognitive practices," which he defines as "manipulations of an external representation to complete a cognitive task" (2010, p. 238).

If PP is right, then the learning process could be optimized by designing environments in order to provide the cycle of action and perception with precisely controlled feedback (prediction error). With the growing commercial availability of immersive virtual reality equipment, educators could design learning environments (or help students design their own environments) without the messy constraints of the physical world. PP may give us a framework with which to understand—and predict—the detailed bodily movements of subjects as they attempt to minimize their own prediction error. Using this framework, we can design systems that would optimize skill acquisition by efficiently predicting the errors that learners will make. This method could be fruitfully applied in the abstract (mathematics), the concrete (skiing), and in-between (foreign languages). Along these lines, the insights of PP, together with emerging technology, can lead to powerful new educational techniques.

Psychology is also applied in some areas of public policy. Clark mentions that PP challenges Kahneman's well-known model of human thinking as consisting of a fast automatic system and a slower deliberative system (p. 18). Kahneman's model has been applied as a basis for influential recommendations about laws and public policy in the United States (Thaler & Sunstein 2008; Sunstein 2014). If PP homes in on a more accurate model of the thinking process, then we ought to use it, rather than (or as a complement to?) the dual systems model as a basis for policy making. Clark's interpretation of PP suggests that we have a highly flexible range of cognitive systems, not limited to Kahneman's two.

For example, one application of Kahneman's model might involve the installation of environmental elements meant to appeal to the fast thinking system, to "nudge" agents towards making decisions in their best interest. If Clark is correct, we might consider even more sophisticated environmental features that have the goal of helping agents to deploy their range of cognitive strategies more efficiently. Clark's ideas of context sensitivity and designer environments are both relevant here. As a society we may wish somehow to create environments and contexts that take advantage of the large repertoire of cognitive strategies available to us, according to Clark's version of PP (see Levy 2012, for example).

The final topic I'd like to mention in this section is what is best described in general terms as social interaction. I mean to indicate a number of related topics here, but the main issue is how PP might relate to the well-known philosophical topic of the way in which we understand and explain our behavior to one an-

other. Recall, for instance, Donald Davidson's (1963) claim that our explanation of our behavior in terms of reasons is a kind of causal explanation—reasons as causes. On his influential view, the connection between reason and actions is a causal connection. In contrast, recall Paul Churchland's envisioning of the golden age of psychology in which we dispose of folk psychological reason-giving in favor of more precise neurophysiological explanations of behavior (1981). According to Churchland's radical alternative, the causes of actions are not reasons as expressed using natural language. Instead, our actions are caused by patterns of neurons firing, patterns that can be described using mathematical tools such as a multidimensional state space. In opposition to Churchland's grand vision, we have Jerry Fodor's claim that the realization of such a vision would be "the greatest intellectual catastrophe in the history of our species" (1987, p. xii). Is PP the beginning of Churchland's grand vision coming to pass? Is a great intellectual catastrophe looming?

On one hand, PP seems like an obvious departure from folk psychology: Try explaining your X-ing to someone by claiming that you X-ed in order to minimize prediction error! One big issue here will be the way in which we think about agency itself. It seems mistaken to say that minimizing prediction error is something done by an agent. Such a process seems to be better described as occurring sub-personally. On the other hand, it is not inconceivable that propositional attitudes can capture the dynamics of prediction error minimization on a suitably coarse-grained level, perhaps along the lines suggested using symbolic dynamics (Dale & Spivey 2005; Atmanspacher & beim Graben 2007; Spivey 2007, Ch. 10). I suggest that these fascinating issues warrant further investigation. In particular, further investigation ought to incorporate Clark's ideas of maximal context sensitivity and the importance of designer environments.

The way in which we understand each other's behavior is also directly relevant for moral responsibility. Following Peter Strawson's seminal "Freedom and Resentment" (1962), philosophers have started thinking about moral responsibility in terms of our reactions to one another, reactions that involve holding each other accountable. On one influential view, we hold each other accountable when our actions issue from our own reasons-responsive mechanisms (Fischer & Ravizza 1998). On a more recent proposal, holding each other accountable is best modeled as a kind of conversation (McKenna 2012). These proposals depend, in important ways, on assumptions about human psychology. In particular, they depend on our practice of giving reasons for behavior. As PP suggests a new fundamental underlying principle of behavior, our practices of holding each other accountable may be approached from a new perspective. The new challenge in this area will be to reconcile (if possible) the practice of giving reasons, on one hand, with PP's account of behavior in terms of error minimization on the other.

5 Conclusion

The main theme of my commentary might appear to be driven by an overexcited optimism for the new theory. To be clear, I have not claimed that PP is correct. Even its main proponents are quick to point out that important open issues remain. My claim is that it is worthwhile to consider the full implications of PP, given the convincing evidence presented so far. In this commentary, I have tried to suggest some of the implications that have not yet been mentioned—implications for perceptual content, consciousness, and applied cognitive neuroscience. These implications can be summarized as follows:

1. PP urges an organism-relative conception of perceptual content.
2. Historical *a priori* accounts of the structure of perceptual experience converge with results from PP.
3. There are a number of areas in which PP can find important practical applications.

The final section includes some challenges for future research. The main challenge is one that

has been familiar in one form or another for several decades in the philosophy of mind. This challenge is to address the tension between the way in which we understand and explain our behavior using natural language, on one hand, and our best theory of human behavior from cognitive neuroscience, which, arguably, is PP, on the other hand. In closing I should note that even if key elements of PP are eventually rejected, it might still turn out that our best model of the mind supports some of the themes I have been discussing.

Acknowledgments

I thank Thomas Metzinger and Jennifer Windt for helpful detailed comments on an earlier draft. This research was supported by the EC Project VERE, funded under the EU 7th Framework Program, Future and Emerging Technologies (Grant 257695).

References

Akins, K. (1996). Of sensory systems and the "aboutness" of mental states. *Journal of Philosophy, 93* (7), 337-372. 10.2307/2941125

Anscombe, G. E. M. (1957). *Intention.* Oxford, UK: Basil Blackwell.

Atmanspacher, H. & beim Graben, P. (2007). Contextual emergence of mental states from neurodynamics. *Chaos and Complexity Letters, 2* (2-3), 151-168.

Bernet, R., Kern, I. & Marbach, E. (1993). *An introduction to Husserlian phenomenology.* Evanston, IL: Northwestern University Press.

Brockman, J. (1996). *Third culture: Beyond the scientific revolution.* New York, NY: Touchstone.

Chalmers, D. (1996). *The conscious mind.* Oxford, UK: Oxford University Press.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78* (2), 67-90. 10.1111/j.1467-9973.1992.tb00550.x

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

——— (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Science, 36* (3), 1-73. 10.1017/S0140525X12000477

——— (2015). Embodied Prediction. In T. Metzinger & J. M. Windt (Eds.) *Open MIND.* Frankfurt a.M., GER: MIND Group.

Dale, R. & Spivey, M. (2005). From apples and oranges to symbolic dynamics: A framework for conciliating notions of cognitive representation. *Journal of Experimental and Theoretical Artificial Intelligence, 17* (4), 317-342. 10.1080/09528130500283766

Davidson, D. (1963). Actions, reasons and causes. *Journal of Philosophy, 60* (23), 685-700. 10.2307/2023177

Dennett, D. (1991). *Consciousness explained.* Boston, MA: Little, Brown, & Co.

Dixon, J., Kelty-Stephen, D. & Anastas, J. (2014). The embodied dynamics of problem solving: New structure from multiscale interactions. In L. Shapiro (Ed.) *The Routledge handbook of embodied cognition.* London, UK: Routledge.

Dretske, F. (1995). *Naturalizing the mind.* Cambridge, MA: MIT Press.

Fischer, J. M. & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility.* Cambridge, UK: Cambridge University Press.

Fodor, J. (1987). *Psychosemantics.* Cambridge, MA: MIT Press.

Freeman, J. & Simoncelli, E. (2011). *Metamers of the ventral stream.* . 10.1038/nn.2889

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, *360* (1456), 815-836. 10.1098/rstb.2005.1622

——— (2008). Hierarchical models in the brain. *PloS Comput Biol*, *4* (11), e1000211. 10.1371/journal.pcbi.1000211

——— (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11* (2), 127-138. 10.1038/nrn2787

——— (2013). Life as we know it. *Journal of the Royal Society Interface*, *10* (86), 1-12. 10.1098/rsif.2013.0475

Friston, K., Kilner, J. & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology - Paris*, *100*, 70-87. 10.1016/j.jphysparis.2006.10.001

Gregory, R. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society B*, *290* (1038), 181-197.

Haynie, D. (2008). *Biological thermodynamics.* Cambridge, UK: Cambridge University Press.

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

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

——— (2014). The self-evidencing brain. *Noûs*, 1-27. 10.1111/nous.12062

Hopp, W. (2011). *Perception and knowledge: A phenomenological account.* Cambridge, UK: Cambridge University Press.

Hume, D. (1993). *An enquiry concerning human understanding.* Indianapolis, IN: Hackett Publishing.

Husserl, E. (1900). *Logische Untersuchungen.* London, UK: Routledge.

——— (1966). *Husserliana XI Analysen zur passiven Synthesis.* Dordrecht, NL: Kluwer.

——— (1969). *Husserliana X zur Phänomenologie des inneren Zeitbewusstseins (1893-1917).* Dordrecht, NL: Kluwer.

——— (1973). *Husserliana XVI Ding und Raum: Vorlesungen 1907.* Dordrecht, NL: Kluwer.

Jackendoff, R. (1987). *Consciousness and the computational mind.* Cambridge, MA: MIT Press.

Jacob, P. (2010). Intentionality. *Stanford Encyclopedia of Philosophy*, *Fall.* http://plato.stanford.edu/entries/intentionality/

Kant, I. (1998). *Kritik der reinen Vernunft.* Hamburg, GER: Meiner Verlag.

Levy, N. (2012). Ecological engineering: Reshaping our environments to achieve our goals. *Philosophy and Technology*, *25* (4), 589-604.

Lewis, D. (1980). Veridical hallucination and prosthetic vision. *Australasian Journal of Philosophy*, *58* (3), 239-249. 10.1080/00048408012341251

Madary, M. (2012a). Husserl on Perceptual Constancy. *European Journal of Philosophy*, *20* (1), 145-165. 10.1111/j.1468-0378.2010.00405.x

——— (2012b). How would the world look if it looked as if it were encoded as an intertwined set of probability density distributions? *Frontiers in Psychology*, *3*. 10.3389/fpsyg.2012.00419

——— (2013). Anticipation and variation in visual content. *Philosophical Studies*, *165*, 335-347. 10.1007/s11098-012-9926-3

Marr, D. (1982). *Vision: A computational approach.* New York, NY: Freeman and Co.

Maturana, H. R. & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living.* Dordrecht, NL: Reidel.

McKenna, M. (2012). *Conversation and responsibility.* Oxford, UK: Oxford University Press.

Menary, R. (2010). Cognitive integration and the extended mind. In R. Menary (Ed.) *The extended mind.* Cambridge, MA: MIT Press.

Merleau-Ponty, M. (1962). *Phenomenology of perception.* London, UK: Routledge.

Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity.* Cambridge, MA: MIT Press.

——— (2009). *The ego tunnel.* New York, NY: Basic Books.

Metzinger, T. & Gallese, V. (2003). The emergence of a shared action ontology: Building blocks for a theory. *Consciousness and Cognition*, *12* (4), 549-571. 10.1016/S1053-8100(03)00072-2

Millikan, R. (1984). *Language, thought and other biological objects.* Cambridge, MA: MIT Press.

——— (1995). Pushmi-pullyu representations. In J. Tomberlin (Ed.) *Philosophical perspectives 9: AI, connectionism, and philosophical psychology.* Atascadero, CA: Ridgeview Publishing Company.

——— (2004). *Varieties of meaning: The 2002 Jean-Nicod Lectures.* Cambridge, MA: MIT Press.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83* (4), 435-450.

Noë, A. (2004). *Action in perception.* Cambridge, MA: MIT Press.

O'Regan, K. & Noë, A. (2001). A sensorimotor approach to vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 939-973.

Prinz, J. (2012). *The conscious brain: How attention engenders experience.* Oxford, UK: Oxford University Press.

Rowlands, M. (2010). *The new science of the mind.* Cambridge, MA: MIT Press.

Searle, J. (1983). *Intentionality: An essay in the philosophy of mind.* Cambridge, MA: Cambridge University Press.

Seth, A. (2014). A predictive processing theory of sensorimotor contingengies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, *5* (2), 97-118. 10.1080/17588928.2013.877880

Snow, C. P. (1959). *The two cultures.* Cambridge, UK: Cambridge University Press.

Spivey, M. (2007). *The continuity of mind.* Oxford, UK: Oxford University Press.

Spratling, M. (2013). Distinguishing theory from implementation in predictive coding accounts of brain function. *Behavioral and Brain Sciences*, *36* (3), 51-52. 10.1017/S0140525X12002178

Strawson, P. F. (1962). Freedom and Resentment. *Proceedings of the British Academy*, *48*, 1-25.

Sunstein, C. (2014). *Why nudge?: The politics of libertarian paternalism.* New Haven, CT: Yale University Press.

Thaler, R. & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* New Haven, CT: Yale University Press.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind.* Cambridge, MA: Harvard University Press.

Tye, M. (2000). *Consciousness, color, and content.* Cambridge, MA: MIT Press.

von Helmholtz, H. (1867). *Handbuch der Physiologischen Optik.* Leipzig, GER: Leopold Voss.

von Uexküll, J. (1934). A stroll through the worlds of animals and men. In K. Lashley (Ed.) *Instinctive behavior.* New York, NY: International Universities Press.

Predicting Peace: The End of the Representation Wars

A Reply to Michael Madary

Andy Clark

Michael Madary's visionary and incisive commentary brings into clear and productive focus some of the deepest, potentially most transformative, implications of the Predictive Processing (PP) framework. A key thread running through the commentary concerns the active and "organism-relative" nature of the inner states underlying perception and action. In this Reply, I pick up this thread, expanding upon some additional features that extend and underline Madary's point. I then ask, What remains of the bedrock picture of inner states bearing familiar representational contents? The answer is not clear-cut. I end by suggesting that we have here moved so far from a once-standard complex of ideas concerning the nature and role of the inner states underlying perception and action that stale old debates concerning the existence, nature, and role of "internal representations" should now be abandoned and peace declared.

Keywords
Action | Action-oriented perception | Content | Enaction | Intentionality | Perception | Perceptual content | Predictive coding | Predictive processing | Representation

Author

Andy Clark
Andy.Clark@ed.ac.uk
University of Edinburgh
Edinburgh, United Kingdom

Commentator

Michael Madary
madary@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Editors

Thomas Metzinger
metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt
jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 Organism-relative content

I'm hugely indebted to Michael Madary for his visionary and incisive commentary. The commentary covers three topics – the nature of perceptual content, the structure of experience, and some practical implications of the PP (Predictive Processing) framework. Each one deserves a full-length paper in reply, but I will restrict these brief comments to the first topic – the nature of perceptual content. Should the PP vision prove correct, Madary suggests, this would transform our understanding of the nature and role of perceptual content, with potential consequences for the larger project of naturalizing mental content. Driving such sweeping and radical reform is (Madary argues) the PP emphasis upon the active contribution of the organism to the generation of perceptual states. There is an active contribution, Madary (this collection, section 2) suggests, insofar as PP depicts perceptual states as "generated internally and spontaneously by the internal dynamics of the generative model" (p. 3).

Such a claim clearly requires careful handling. For even the most staunchly feedforward model of perception requires a substantial contribution from the organism. It is thus the nature, not the existence, of that contribution

Clark, A. (2015). Predicting Peace: The End of the Representation Wars-A Reply to Michael Madary.
In T. Metzinger & J. M. Windt (Eds). *Open MIND:* 7(R). Frankfurt am Main: MIND Group. doi: 10.15502/9783958570979

that must make the difference. Elaborating upon this, Madary notes that ongoing endogenous activity plays a leading role in the PP story. One might say: the organism's generative model (more on which later) is already active, attempting to predict the incoming sensory flow. The flow of incoming information is thus rapidly flipped into a flow tracking "unexpected salient deviation". Identical inputs may thus result in very different perceptual states as predictions alter and evolve. An important consequence, highlighted by Madary, is that different histories of interaction will thus result in different perceptual contents being computed for the very same inputs. Different species, different niches, differences of bodily form, and differences of proximal goals and of personal history are all thus apt (to varying degrees) to transform what is being predicted, and hence the contents properly delivered by the perceptual process.

Those contents are further transformed by a second feature of the PP account: the active selection of perceptual inputs. For at the most fundamental level, the PP story does not depict perception as a process of building a representation of the external world at all. Instead, it depicts perception as just one part of a cohesive strategy for keeping an organism within a kind of "window of viability". To this end the active organism both predicts *and selects* the evolving sensory flow, moving its body and sensory organs so as to expose itself to the sensory stimulations that it predicts. In this way, some of our predictions act as self-fulfilling prophecies, enabling us to harvest the predicted sensory streams. These two features (endogenous activity and the self-selection of the sensory flow) place PP just about maximally distant from traditional, passive "feedforward hierarchy" stories. They are rather (as Mike Anderson once commented to me) the ultimate expression of the "active perception" program.

Here too, though, we should be careful to nuance our story correctly. For part of maintaining ourselves in a long-term window of viability may involve not just seeking out the sensory flows we predict, but the active elicitation of many that we don't! PP may, in fact, mandate all manner of short-term explorations and self-destabilizations. But such delicacies (though critically important- see Clark (in press) chapters 8 and 9) may safely be left for another day. The present upshot (Madary this collection, section 2) is simply that PP, instead of depicting perception as a mechanism for revealing "what is where" in the external world, turns out to be a mechanism for engaging the external world in ways that say as much about the organism (and its own history) as they do about the world outside. To naturalize intentionality, then, "all" we need do is display the mechanisms by which such ongoing viability-preserving engagements are enabled, and make intelligible that such mechanisms can deliver the rich and varied grip upon the world that we humans enjoy. This, of course, is exactly what PP sets out to achieve.

2 Structural coupling and the bringing forth of worlds[1]

Madary notes, more or less in passing, that the PP vision of "organism-relative perceptual content" bears a close resemblance to views that have been defended under the broad banner of "enactivism". I want to pick up on this hint, and suggest that the PP account actually sets the scene for peace to be declared between the once-warring camps of representationalism and enactivism. Thus consider the mysterious-sounding notion of "enacting a world", as that notion appears in Varela et al. (1991)[2]. Varela et al. write that:

> The overall concern of an enactive approach to perception is not to determine how some perceiver-independent world is to be recovered; it is, rather, to determine the common principles or lawful linkages between sensory and motor systems that explain how

1 Parts of this section condense and draw upon materials from Clark (in press).

2 There is now a large, and not altogether unified, literature on enaction. For our purposes, however, it will suffice to consider only the classic statement by Varela et al. (1991). Important contributions to the larger space of enactivist, and enactivist-inspired, theorizing include Noë (2004, 2010, this collection), Thompson (2010), and Froese & Di Paolo (2011). The edited volume by Stewart et al. (2010) provides an excellent window onto much of this larger space.

> action can be perceptually-guided in a perceiver-dependent world. (1991, p. 173)

This kind of relation is described by Varela et al. as one of "structural coupling" in which "the species brings forth and specifies its own domain of problems" (1991, p. 198) and in that sense "enacts" or brings forth (1991, p. 205) its own world. In discussing these matters, Varela et al. are also concerned to stress that the relevant histories of structural coupling may select what they describe as "non-optimal" features, traits, and behaviors: ones that involve "satisficing" (see Simon 1956) where that means settling for whatever "good enough" solution or structure "has sufficient integrity to persist" (Varela et al. 1991, p. 196). PP, I will now suggest, has the resources to cash these enactivist cheques, depicting the organism and the organism-salient world as bound together in a process of mutual specification in which the simplest approximations apt to support a history of viable interaction are the ones that are learnt, selected, and maintained.

The simplest way in which a PP-style organism might be said to actively construct its world is by sampling. Action, as Madary noted, serves perception by moving the body and sense-organs around in ways that aim to "serve up" predicted sequences of high-reliability, task-relevant information. In this way, different organisms and individuals may selectively sample in ways that both actively construct and continuously confirm the existence of different "worlds". It is in this sense that, as Friston, Adams, and Montague (Friston et al. 2012, p. 22) comment, our implicit and explicit models might be said to "create their own data".[3] Furthermore, the PP framework depicts perception and action as a single (neurally distributed) process whose goal is the reduction of salient prediction-error. To be sure, "sensory" and "motor" systems specialize in different predictions. But the old image of sensory information IN and motor output OUT is here abandoned. Instead, there is a unified sensorimotor system aiming to predict the full range of sensory inputs – inputs that are often at least partially self-selected and that include exteroceptive, proprioceptive (action-determining), and interoceptive elements. Nor is it just the sensorimotor system that is here in play. Instead, the whole embodied organism (as Madary notes) is treated as a prediction-error minimizing device.

The task of the generative model in all these settings is (as noted in Clark this collection) to capture the simplest approximations that will support the actions required to do the job. And that means taking into account whatever work can be done by a creature's morphology, physical actions, and socio-technological surroundings. Such approximations are constrained to "provide the simplest (most parsimonious) explanations for sampled outcomes" (Friston et al. 2012, p. 22). This respects the enactivist's stress on biological frugality, satisficing, and the ubiquity of simple but adequate solutions that make the most of brain, body, and world. At this point, all the positive enactivist cheques mentioned above have been cashed.

But one outstanding debt remains. To broker real and lasting peace, we must tiptoe bravely back into some muddy and contested territory: the smoking battleground of the Representation wars.

3 Representations: What are they good for?

PP, Madary suggests, provides a new kind of lever for naturalizing intentionality and mental content. Might it also offer a new perspective upon the vexed topic of internal representation? Varela et al. are explicit that, on the enactivist conception "cognition is no longer seen as problem solving on the basis of representations"

3 Such a process repeats at several organizational scales. Thus we humans do not merely sample some natural environment. We also structure that environment by building material artifacts (from homes to highways), creating cultural practices and institutions, and trading in all manner of symbolic and notational props, aids, and scaffoldings. Some of our practices and institutions are also designed to *train us to sample* our human-built environment more effectively – examples would include sports practice, training in the use of specific tools and software, learning to speed-read, and many, many more. Finally, some of our technological infrastructure is now self-altering in ways that are designed to reduce the load on the predictive agent, learning from our past behaviors and searches so as to serve up the right options at the right time. In all these ways, and at all these interacting scales of space and time, we build and selectively sample the very worlds that - in iterated bouts of statistically-sensitive interaction - install the generative models that we bring to bear upon them.

(1991, p. 205). PP, however, deals extensively in internal models – models that may (see Clark this collection) be rich, frugal, and all points in-between. The role of such models is to control action by predicting and bringing about complex plays of sensory data. This, the enactivist might fear, is where our promising story about neural processing goes conceptually astray. Why not simply ditch the talk of inner models and internal representations and stay on the true path of enactivist virtue?

This issue requires a lot more discussion than I can attempt here.[4] Nonetheless, the remaining distance between PP and the enactivist may not be as great as that bald opposition suggests. We can begin by reminding ourselves that PP, although it openly trades in talk of inner models and representations, invokes representations that are action-oriented through and through. These are representations that are fundamentally in the business of serving up actions within the context of rolling sensorimotor cycles. Such representations aim to *engage* the world, rather than to depict it in some action-neutral fashion, and they are firmly rooted in the history of organism-environment interactions that served up the sensory stimulations that installed the probabilistic generative model. What is on offer is thus just about maximally distant from a passive ("mirror of nature" – see Rorty 1979) story about the possible fit between model and world. For the test of a good model is how well it enables the organism to engage the world in a rolling cycle of actions that maintain it within a window of viability. The better the engagements, the lower the information-theoretic free energy (this is intuitive, since more of the system's resources are being put to "effective work" in engaging the world). Prediction error reports this information-theoretic free energy, which is mathematically constructed so as always to be greater than "surprisal" (where this names the sub-personally computed implausibility of some sensory state given a model of the world – see Tribus 1961). Notice also that the prediction task uses only information *clearly available to the organism*, and is ultimately defined over the energies that impinge on the organism's sensory surfaces. But finding the best ways to predict those energetic impacts can (as substantial bodies of work in machine learning amply demonstrate[5]) yield a structured grip upon a world of interacting causes.

This notion of a *structured* grip is important. Early connectionist networks were famously challenged (Fodor & Pylyshyn 1988) by the need to deal with structure – they were unable to capture part-whole hierarchies, or complex nested structures in which larger wholes embed smaller components, each of which may itself be some kind of structured entity. For example, a city scene may consist of a street populated by shops and cars and people, each of which is also a structured whole in its own right. Classical approaches benefitted from an easy way of dealing with such issues. There, digital objects (symbol strings) could be composed of other symbols, and equipped with pointers to further bodies of information. This apparatus was (and remains) extremely biologically suspect, but it enabled nesting, sharing, and recombination on a grand scale – see Hinton (1990) for discussion. Such systems could easily capture structured (nested, often hierarchical) relationships in a manner that allowed for easy sharing and recombination of elements. But they proved brittle and inflexible in other ways, failing to display fluid context-sensitive responsiveness, and floundering when required to guide behavior in time-pressured real-world settings.[6]

Connectionist research has since spawned a variety of methods – some more successful than others - for dealing with structure in various domains. At the same time, work in robotics and in embodied and situated cognitive science has explored the many ways in which structure in the environment (including the highly structured artificial environments of text and external symbol systems) could be exploited so as to reap some of the benefits associated with classical forms of in-

4 I have engaged such arguments at length elsewhere – see Clark (1989, 1997, 2008, 2012). For sustained arguments *against* the explanatory appeal to internal representation, see Ramsey (2007), Chemero (2009), Hutto & Myin (2013). For some useful discussion, see Sprevak (2010, 2013), Gallagher et al. (2013).

5 For reviews and discussions, see Bengio (2009), Huang & Rao (2011), Hinton (2007), and Clark (in press).

6 For a sustained discussion of these failings, and the attractions of connectionist (and post-connectionist) alternatives, see Clark (1989, 1993, 2014), Bechtel & Abrahamsen (2002), Pfeifer & Bongard (2007).

ner encoding, without (it was hoped) the associated costs of biological implausibility – see, for example, Pfeifer & Bongard (2007). Perhaps the combination of a few technical patches and a much richer reliance upon the use of structured external resources would address the worries about dealing with structure? Such was the hope of many, myself included.

On this project, the jury is still out. But PP can embrace these insights and economies while providing a more powerful overall solution. For it offers a biologically plausible means, consistent (we saw) with as much reliance on external scaffolding as possible, of internally encoding and deploying richly structured bodies of information. This is because each PP level (perhaps these correspond to cortical columns – this is an open question) treats activity at the level below as if it were sensory data, and learns compressed methods to predict those unfolding patterns. This results in a very natural extraction of nested structure in the causes of the input signal, as different levels are progressively exposed to different recodings, and re-re-codings of the original sensory information. These re-re-codings (I think of them as representational re-descriptions in much the sense of Karmiloff-Smith 1992) enable us, as agents, to lock us onto worldly causes that are ever more recondite, capturing regularities visible only in patterns spread far in space and time. Patterns such as weather fronts, persons, elections, marriages, promises, and soccer games. Such patterns are the stuff of which human lives, and human mental lives, are made. What locks the *agent* on to these familiar patterns is, however, the whole multi--level processing device (sometimes, it is the whole machine in action). That machine works (if PP is correct) because each level is driven to try to find a compressed way to predict activity at the level below, all the way out to the sensory peripheries. These nested compressions, discovered and annealed in the furnace of action, are what I (following Hinton 1990) would like to call "internal representations".

What are the *contents* of the many states governed by the resulting structured, multi-level, action-oriented probabilistic generative models? The generative model issues predictions that estimate various identifiable worldly states (including states of the body, and the mental states of other agents).[7] But it is also necessary, as we saw in Clark (this collection) to estimate the context-variable reliability (precision) of the neural estimations themselves. It is these precision-weighted estimates that drive action, and it is action that then samples the scene, delivering percepts that select more actions. Such looping complexities exacerbate an important consequence that Madary nicely notes. They make it even harder (perhaps impossible) adequately to capture the contents or the cognitive roles of many key inner states and processes using the terms and vocabulary of ordinary daily speech. That vocabulary is "designed" for communication, and (perhaps) for various forms of cognitive self-stimulation (see Clark 2008). The probabilistic generative model, by contrast, is designed to engage the world in rolling, uncertainty-modulated, cycles of perception and action. Nonetheless, high-level states of the generative model will target large-scale, increasingly invariant patterns in space and time, corresponding to (and allowing us to keep track of) specific individuals, properties, and events despite large moment-by-moment variations in the stream of sensory stimulation. Unpacked via cascades of descending prediction, such higher-level states simultaneously inform both perception and action, locking them into continuous circular causal flows. Instead of simply describing "how the world is", these models - even when considered at those "higher" more abstract levels - are geared to engaging those aspects of the world that matter to us. They are delivering a grip on the *patterns that matter* for the *interactions that matter*.

Could we perhaps (especially given the likely difficulties in specifying intermediate-level contents in natural-language terms) have told our story in entirely non-representational terms, without invoking the concept of a hierarchical probabilistic generative *model* at all? One should always beware of sweeping assertions about what might, one day, be explanatorily possible! But as things stand, I simply don't see how this is to be achieved. For it is surely that very model-invoking

7 Bayesian perceptual and sensorimotor psychology (see for example, Rescorla 2013; Körding & Wolpert 2006) already has much to say about just what worldly and bodily states these may be.

schema that allows us to understand how it is that these looping dynamical regimes arise and enable such spectacular results. The regimes arise and succeed because the system self-organizes around prediction-error so as to capture organism-salient patterns, at various scales of space and time, in the (partially self-created) input stream. These patterns specify complex, inter-animated structures of bodily and worldly causes. Subtract this guiding vision and what remains is just the picture of complex looping dynamics spanning brain, body, and world. Consider those same looping dynamics from the multi-level model-invoking explanatory perspective afforded by PP, however, and many things fall naturally into place. We see how statistically-driven learning can unearth interacting distal and bodily causes in the first place, revealing a structured world of human-sized opportunities for action; we see why, and exactly how, perception and action can be co-constructed and co-determining; and we unravel the precise (and happily un-mysterious) sense in which organisms may be said to bring forth their worlds.

4 Predicting peace: An end to the war over internal representation

Dynamically speaking, the whole embodied, active system here self-organizes around the organismically-computable quantity "prediction error". This is what delivers that multi-level, multi-area, grip on the evolving sensory barrage – a grip that must span multiple spatial and temporal scales. Such a grip simultaneously determines perception and action, and it selects (enacts) the ongoing stream of sensory bombardment itself. The generative model that here issues sensory predictions is thus nothing but that multi-level, multi-area[8], multi-scale, body-and-action involving grip on the unfolding sensory stream. To achieve that grip is to know the structured and meaningful world that we encounter in experience and action.

Is this an inner economy bloated with representations, detached from the world? Not at all. This is an inner economy geared for action, whose inner states bear contents in virtue of the way they lock embodied agents onto properties and features of their worlds. But it is simultaneously a structured economy built of nested systems, whose communal project is both to model and engage the (organism-relative) world.

8 The point about multiple areas (not just multiple levels within areas) is important, but it is often overlooked in philosophical discussions of predictive processing. Different neural areas are best-suited – by location, inputs, structure, and/or cell-type - to different kinds of prediction. So the same overarching PP strategy will yield a complex economy in which higher-levels predict lower levels, but different areas learn to trade in very different kinds of prediction. This adds great dynamical complexity to the picture, and requires some means for sculpting the flow of information among areas. I touch on these issue in Clark (this collection). But for a much fuller exploration, see Clark (in press).

References

Bechtel, W. & Abrahamsen, A. (2002). *Connectionism and the mind: Parallel processing, dynamics, and evolution in networks.* Oxford, UK: Basil Blackwell.

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, *2* (1), 1-127. 10.1561/2200000006

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

Clark, A. (1989). *Microcognition: Philosophy, cognitive science and parallel distributed processing.* Cambridge, MA: MIT Press.

——— (1993). *Associative engines: Connectionism, concepts and representational change.* Cambridge, MA: MIT Press.

——— (1997). *Being there: Putting brain, body and world together again.* Cambridge, MA: MIT Press.

——— (2008). *Supersizing the mind: Action, embodiment, and cognitive extension.* New York, NY: Oxford University Press.

——— (2012). Dreaming the whole cat: Generative models, predictive processing, and the enactivist conception of perceptual experience. *Mind*, *121* (483), 753-771. 10.1093/mind/fzs106.

——— (2014). *Mindware: An introduction to the philosophy of cognitive science.* New York, NY: Oxford University Press.

——— (2015). Embodied prediction. In T. Metzinger & J. M. Windt (Eds.) *Open MIND.* Frankfurt a. M., GER: MIND Group.

Clark, A. (in press). *Surfing uncertainty: Prediction, action, and the embodied mind.* New York, NY: Oxford University Press.

Fodor, J. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28* (1-2), 3-71. 10.1016/0010-0277(88)90031-5

Friston, K., Adams, R. & Montague, R. (2012). What is value—Accumulated reward or evidence? *Frontiers in Neurorobotics*, *6*. 10.3389/fnbot.2012.00011

Froese, T. & Di Paolo, E. A. (2011). The enactive approach: Theoretical sketches from cell to society. *Pragmatics and Cognition*, *19* (1), 1-36. 10.1075/pc.19.1.01fro

Gallagher, S., Hutto, D., Slaby, J. & Cole, J. (2013). The brain as part of an enactive system. *Behavioral and Brain Sciences*, *36* (4), 421-422. 10.1017/S0140525X12002105

Hinton, G. E. (1990). Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence*, *46* (1-2), 47-75. 10.1016/0004-3702(90)90004-J

——— (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, *11* (10), 428-434. 10.1016/j.tics.2007.09.004

Huang, Y. & Rao, R. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2* (5), 580-593. 10.1002/wcs.142

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

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science.* Cambridge, MA: MIT Press/Bradford Books.

Körding, K. & Wolpert, D. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10* (7), 319-326. 0.1016/j.tics.2006.05.003

Madary, M. (2015). Extending the explanandum for predictive processing - A commentary on Andy Clark. In T. Metzinger & J. M. Windt (Eds.) *Open MIND.* Frankfurt a. M., GER: MIND Group.

Noë, A. (2004). *Action in perception.* Cambridge, MA: MIT Press.

——— (2010). *Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness.* New York, NY: Farrar, Straus and Giroux.

——— (2015). Concept pluralism, direct perception, and the fragility of presence. In T. Metzinger & J. M. Windt (Eds.) *Open MIND.* Frankfurt a. M., GER: MIND Group.

Pfeifer, R. & Bongard, J. (2007). *How the body shapes the way we think: A new view of intelligence.* Cambridge, MA: MIT Press.

Ramsey, W. M. (2007). *Representation reconsidered.* Cambridge, UK: Cambridge University Press.

Rescorla, M. (2013). Bayesian perceptual psychology. *Oxford handbook of the philosophy of perception (forthcoming).* Oxford, UK: Oxford University Press.

Rorty, R. (1979). *Philosophy and the mirror of nature.* Princeton, NJ: Princeton University Press.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63* (2), 129-138. 10.1037/h0042769

Sprevak, M. (2010). Computation, individuation, and the received view on representation. *Studies in History and Philosophy of Science*, *41* (3), 260-270. 10.1016/j.shpsa.2010.07.008

——— (2013). Fictionalism about neural representations. *The Monist*, *96* (4), 539-560. 10.5840/monist201396425

Stewart, J., Gapenne, O. & Di Paolo, E. (Eds.) (2010). *Enaction: Towards a new paradigm for cognitive science.* Cambridge, MA: MIT Press.

Thompson, E. (2010). *Mind in life: Biology, phenomenology, and the sciences of mind.* Cambridge, MA: Harvard University Press.

Tribus, M. (1961). *Thermodynamics and thermostatics: An introduction to energy, information and states of matter, with engineering applications.* New York, NY: D. Van Nostrand Company Inc.

Varela, F., Thompson, E. & Rosch, E. (1991). *The embodied mind.* Cambridge, MA: MIT Press.