

**Andrew Buskell**

## **Book review: Andy Clark // surfing uncertainty: prediction, action, and the embodied brain**

**Article (Published version)  
(Non-refereed)**

**Original citation:**

Buskell, Andrew (2017) *Book review: Andy Clark // surfing uncertainty: prediction, action, and the embodied brain*. *British Journal for the Philosophy of Science - Review of Books* . ISSN 0007-0882


© 2017 The Author

This version available at: <http://eprints.lse.ac.uk/73659/>

Available in LSE Research Online: April 2017

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

# Andy Clark // Surfing Uncertainty: Prediction, Action, and the Embodied Brain

 [bjpsbooks.wordpress.com/2017/01/31/andy-clark-surfing-uncertainty-prediction-action-and-the-embodied-brain/](https://bjpsbooks.wordpress.com/2017/01/31/andy-clark-surfing-uncertainty-prediction-action-and-the-embodied-brain/)

View all posts by bjpsbooks

## Reviewed by Andrew Buskell

### *Surfing Uncertainty: Prediction, Action, and the Embodied Brain*

Andy Clark

Oxford: Oxford University Press, 2016, £20 (hardback)

ISBN 9780190217013

I have a vivid memory from my undergraduate degree in psychology: a short video, no longer than five seconds long, showing the characteristic ‘pinwheel’ activation of orientation-sensitive cells in the primary visual cortex. These cells (clumped into ‘columns’) respond differentially to the orientation of light in the visual field. In the video I watched, this responsiveness was rendered in dramatic fashion: responding to a clock-wise rotating stimuli, the activation pattern of the cells rotated in near perfect synchrony.

I suspect this video is typical of those shown to undergraduates. Typical too, I would wager, is the philosophical picture wedded to it: perception involves the detection of external stimuli and the successive processing and enrichment of that stimuli in service of producing a rich internal representation of the outside world.

According to Andy Clark, this philosophical picture is misleading. In *Surfing Uncertainty: Prediction, Action, and the Embodied Brain*, Clark aims to replace this picture of detection, enrichment, and rich internal representation. His alternative picture is one based upon generative models, precision-weighting, and the detection of prediction-error. While the traditional picture is bottom-up and employs a strategy of building a model of the world that lags *just behind* the world in the here-and-now, Clark’s alternative is almost the complete inverse; it holds that cognition involves top-down models that make guesses as to how the states of the world are likely to unfold, and these predictions guide the (generally, sub-personal) selection of percept, sensory orientation, and action.

I think Clark is right to identify the picture presented in the book as a ‘mid-level organizational sketch’ (p. 2). It provides a general and abstract mechanistic characterisation of the function and organization of the nervous system. The broad contours of this sketch hold that the brain is in the business of acquiring and tuning a model that generates predictions about the temporally fluctuating distal causes of sensory stimulation. Filling in this organizational sketch is the work of the first third of the book—applying it to the wide range of psychological and philosophical issues in psychology, philosophy of mind, and beyond occupies the latter two thirds.

Part one provides the reader with more detail on Clark’s ‘organizational sketch’; what he comes to call hierarchical action-oriented predictive processing (hereafter, PP). As stated above, PP is a characterization of the organization and function of the nervous system. It holds that cognitive systems can be organized into hierarchical levels, and that higher levels of the hierarchy are in the business of predicting the incoming stimuli from lower levels. Errors resulting from mismatches between stimuli and prediction are fed back to the higher levels to revise future guesses. This view, with roots in the work of Karl Friston, Geoffrey Hinton, David Knill, and Alexandre Pouget, turns the traditional picture on its head. Instead of a rich internal representation of the world that must be continually updated, Clark’s picture posits that the nervous system realizes a ‘generative model’, one that produces hypotheses about the spatiotemporal structure of sensory input. This generative model in turn is the result of an extended process of Bayesian inference, where initial guesses as to how the world is organized are inferred on-the-go as individuals develop, and which are tuned and tweaked as individuals engage in the dynamic flow of the world.

If I understand Clark’s account correctly, what agents actually experience are the hypotheses produced by the

generative model. Yet despite the fact that experiences result or are realized by inner structures, one should not think of this generative model in terms of representations that ‘mirror’ the world. This would miss the inherent *strangeness* of the generative model as compared to traditional, folk characterizations of representational content. To give something of a caricature, a folk characterization holds that early perceptual processes are involved in something like feature placing, with representational content increasing in detail and richness until something like a representational duplicate of the world has been achieved, replete with facsimiles of the entities and properties out there in the world. On Clark’s account, however, movement up the representational hierarchy brings with it increasing temporal and spatial scope over sensory stimuli. At the higher levels of this hierarchy, representations consist in probability distributions over temporally extended trajectories of sensory stimulation. This is a radical departure from a picture of representational content that corresponds to or mirrors states of affairs of the world.

Chapter two adds in a crucial component of Clark’s project: precision-weighting. For Clark, ‘we are constantly engaged in attempts to predict precision, that is, to predict the context-varying reliability of our own sensory prediction error’ (p. 58). Developed with increasing precision throughout the book (notably in Chapters 4, 5, and 7) the account of precision-weighting the reader is given is one where variable uncertainty leads to savvy sampling of information-rich patches of the world, particularly those that can help to determine the context that one is in, or the right action to take. In Clark’s terminology, precision-weighting modulates the influence of top-down priors relative to bottom-up stimuli. Normal (that is, non-pathological) cognitive and behavioural engagement with the world comes to rest on precision-weighting achieving the right kind of ‘balance’ between predictions and stimuli.

Despite Clark grouping Chapter 3 in with the first part of the book, I think it sits more easily with the second, thematic part of the book that extends to Chapter 7. In part two, Clark interrogates clumps (there might be no better word here) of psychological, neuroscientific, psychopathological, and cognitive phenomena. These clumps are often broad, involving surveys over a huge range of material and subject matters. In discussing each of these clumps, Clark argues cogently and forcefully that the PP framework provides a powerful means of getting at the empirical facts of the matter.

Because of the large and disparate amount of material covered in Chapters 3 through to 7, it would be foolish to try give a full description of what is covered therein, let alone a satisfactory analysis. Instead, I here summarize the general phenomena discussed in each chapter, which should give the reader some idea of what is in store when they pick up the book. Chapter 3 discusses a ‘cognitive package deal’, with PP taken to give a neat and tidy account as to how episodic memory, imagination, and dreaming could emerge together. Chapter 4 looks at the integration of PP with action, planning, and the control of behaviour. Chapter 5 examines the nature of mirror neurons, proprioceptive information, off-line rehearsal of action, and the role of context and context-detection. Chapter 6 is an examination of the debate between empiricist and nativists about cognitive architecture, and looks at the way in which a PP architecture *qua* an empirical Bayes approach can illuminate these debates, as well as the nature and selection of environmental affordances. Finally, Chapter 7 reflects on the nature of neuropathologies such as schizophrenia, autism, and depersonalization disorder. As stated above, in each of these chapters, Clark labours to show that the PP framework not only captures the phenomena discussed in the literature, but makes new predictions, solves old problems, and ultimately provides a more empirically fruitful means of getting at cognitive phenomena.

Looking at the wide range of material discussed, and the kind of argument that Clark is attempting, it should come as no surprise to note that the most frequently used phrase in the book is ‘*The upshot is...*’. Indeed, to my mind, Chapters 3 through to 7 can be seen as providing an extended inference to the best explanation, with Clark articulating the empirical benefits that come along when one adopts the PP framework. Key among these benefits is the way in which the PP framework can provide a unifying and synthesizing framework for understanding the nervous system and cognitive phenomena.

The last two chapters make up the final part of the book. While there is a great deal of empirical content discussed in these chapters, it seems to me that they are more accurately characterized as providing a picture of human cognition and behaviour in all its richness—what Peter Godfrey-Smith has called a ‘philosophy of nature’. In Chapter 8, Clark considers the ‘dynamic instability’ of human hierarchical PP as a means of capturing

phenomena such as curiosity and neophilia, as well as dissolving worries that the PP will lead one to hide in a 'darkened room'. The last chapter of the book, Chapter 9, situates the cognitive agent in the richness of the wider sociocultural milieu, bringing in concepts of self-organization, creativity, scaffolding, niche-construction, language, and joint-organization to show how PP can be integrated with contemporary work in cultural evolution, enactivism, and embodied cognition.

What I hope comes through from the reflections above is a recognition of the argumentative strategy of the book. The argument is cumulative: part one introduces the pieces of machinery that are deployed in part two and three. In these later parts, similar arguments and strategies are deployed again and again to erode support for the 'enrichment', or 'bottom-up' conception of mentality and cognition, and to cement the empirical merits of the PP picture. By the end of part three, we have been provided with a comprehensive and holistic picture, one that can accommodate, relate, and explain a wealth of cognitive, neuroscientific, and social phenomena.

As is always the case with scholarly monographs, individuals will have wished that some issues were discussed in more detail, or analysed in greater depth. For what it is worth, I would have liked to have further discussion of the 'hierarchy' in 'hierarchical action-oriented predictive processing', as we seem to have been given very different characterisations of what this hierarchy consists in.

Above, I characterized cognitive processing—'ascending the representational hierarchy'—in terms of representations or predictions of increasing spatiotemporal scope. This is a characterization endorsed several times throughout the book (pp. 111, 158, 184). But as the book progresses, Clark introduces alternative characterizations that bring in distinct elements. Notably, in Chapter 5, Clark puts the emphasis on context, and suggests that as representations move up in the representational hierarchy, they represent increasingly sketchy, but high-level, 'gists' of scenarios. This hierarchy tops-off when one reaches a background 'mindset', which provides the 'baseline sets of expectations that are themselves already active when we process even the roughest, coarsest sets of sensory cues from the external (or indeed the internal) environment' (p. 166). Further, in later chapters of the book, we are introduced to the intriguing notion of a 'hyperprior' (pp. 174–5, 188), which the reader is told embodies 'systemic expectations concerning very abstract (at times almost "Kantian") features of the world' (p. 174). Clark uses this characterization (at least) to make sense of the way in which, despite the probabilistic nature of our generative model, we always see the world as being in one determinate state or another.

To my mind, these three characterizations describe very distinct kinds of models, and different pictures as to what it means to 'move up' the representational hierarchy. Spatiotemporal scope is not (necessarily) the same as gist or context, nor are these (necessarily) the same as invariant features of experience or behaviour. Perhaps these features of high-level representations are compatible, or maybe even partially overlapping—but this need not be the case.

I suspect this represents a small if non-trivial conceptual worry for the PP account Clark develops. A more pressing issue concerns those that will deny at the outset the kind of picture that Clark attempts to put forward. Those who are committed to something like 'bottom-up' approaches to cognition will no doubt be able to craft clever hypotheses in order to save the phenomena. Clark even considers several means by which such crafty hypotheses have already been deployed to explain tricky aspects of action control (notably, efference-copy accounts, covered in Chapter 4).

One useful means of organizing some of the arguments of Clark's book, then, might be to bring in what Gerald Holton ([1975], p. 330) called 'themata'. For Holton, such themata represent 'early, unshakable commitments' to alternate conceptual frameworks. Despite the fact that such frameworks appeal to antithetical 'thematic concepts' (Holton's examples, from physics, include dyads like 'continuum' and 'discreteness'), there is plenty of scope for empirical disagreement. As a matter of fact, as Holton notes, it is often the case that in 'the scientists' own public presentation of their work, and during any ensuing scientific controversy, these [thematic] elements are usually not explicitly at issue'. Indeed, lest scientists risk being called philosophers, the focus remains on the adequacy of theories to accommodate and explain empirical phenomena.

It is useful to have Holton's themata in the back of one's mind when reading Clark's book. For while the

overarching message is that it is helpful to switch one's early, unshakeable commitments—from a 'bottom-up' picture of enrichment to a 'top-down' picture of hierarchical predictive processing—much of the substance of the book is an engagement with cutting-edge empirical research from a huge variety of sources. And this makes sense. Themata, as Holton characterizes them, are not hypotheses that can be falsified. They are something much more fundamental, representing perspectives or frames as to how the world is organized. Nonetheless, for Clark, this 'top-down' picture seems to provide a greater and more fruitful empirical grasp on cognitive and neuroscientific phenomena, one that brings with it a comprehensive and unifying perspective on what the brain is doing.

The focus on empirical details is effective in showing the wide-ranging scope and predictive power of Clark's account—but it is likely to leave as many questions unanswered as answered. Proponents of various positions within experimental psychology, neuroscience, and embodied cognition (to name just a few) are likely to feel that issues around mirror neurons, context, skills, and imitation are treated far too hastily, and that this might be reason enough to reject Clark's picture. Philosophers might also think that important issues are not adequately discussed. Mentions of folk psychology, disjunctivism, realism, and the nature of representation are brief, and often find their most interesting arguments hidden away in the book's many footnotes. As a result, philosophers too might question whether we've really been given a 'knock-down' argument in favour of PP.

Such knee-jerk reactions would be unfortunate. Clark's book should not be taken as a definitive account of all the phenomena he discusses, and it will always be the case in a book like this that researchers will feel that their particular research topics could have been given more space and more comprehensive discussion.

Let me conclude with two final remarks. The first is that it is important to recognize the scope and audacity of Clark's book. It synthesizes an absolutely massive amount of material—both empirical and philosophical—and manages to present this material in an illuminating, coherent, and compelling manner. Of course, individuals will disagree on the extent to which this synthesizing vision is successful, as well as on the details of the picture. This brings me to my second point. Like some of his other books (notably, the classic *Microcognition* ([1989]) and the well-known *Being There* ([1996])), this is an agenda-setter. It sets the stage for debates to come. The all-too-brief discussion of topics with massive primary and secondary literatures should be seen as promissory notes. I suspect that Clark, and other philosophers and empirical researchers, will be busy making good on these promissory notes for the foreseeable future.

The upshot of all of this is, to my mind, that this is a truly important book. It is evocatively written and reflects a truly gargantuan amount of work. It sets the stage for future debates not only about the empirical merits of Bayesian characterizations of human cognition, but also the broader philosophical picture in which such Bayesian characterizations are embedded. I predict that many of us will be reading, discussing, and analysing this book in the months and years to come.

*Andrew Buskell*  
*Department of History and Philosophy of Science*  
*University of Cambridge*  
*ab2086@cam.ac.uk*

## **Acknowledgements**

I am exceptionally grateful to both Andy Clark and Dan Williams for their helpful comments on earlier drafts of this review.

## **References**

Clark, A. [1989]: *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*, Cambridge, MA: MIT Press

Clark, A. [1996]: *Being There: Putting Brain, Body, and World Together Again*, Cambridge, MA: MIT Press.

Holton, G. [1975]: 'On the Role of Themata in Scientific Thought', *Science*, **188**, pp. 328–34.