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 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.19.1.01-fro

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.