Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 24 (2018) 17–20

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Comment

The foundation: Mechanism, prediction, and falsification in Bayesian enactivism
Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al.

Micah Allen [a,b,*]

[a] *Institute of Cognitive Neuroscience, UCL, United Kingdom*
[b] *Wellcome Centre for Human Neuroimaging, UCL, United Kingdom*

Received 9 January 2018; accepted 18 January 2018
Available online 2 February 2018
Communicated by J. Fontanari

In Isaac Asimov's science fiction classic, *Foundation*, fictional mathematician Hari Seldon applies his theory of *psychohistory*, a synthesis of psychology, history, and statistical physics, to predict that humanity will suffer a dark age lasting thirty millennia [1]. Although Seldon's psychohistory successfully predicts the future of human society, its basis in the physical law of mass action carries a limitation – it can only do so for sufficiently massive populations (i.e., billions of individuals), rendering it inert at an individual level. This limitation is of course a key source of dramatic tension in the series, in which the individual characters of Asimov's universe grapple with the challenges inherent to applying a lawlike theory of collective action to the constitutive individuals. To avert crisis, Seldon ultimately assembles the namesake Foundation, an interdisciplinary, intergalactic research centre bringing together various biological, physical, and social scientists who ultimately attempt to alter the predicted course of history.

It seems that even in science, life imitates art. In their ambitious new proposal, Ramstead et al. [2] seemingly overcome the limitations of psychohistory, linking the dynamical interaction of cells, brains, persons, and even society itself through the mathematical lens of the free energy principle (FEP). By expanding the FEP to define a new "variational neuroethology" (VNE) of biological life, Ramstead et al. propose to "... to explain and predict how living systems, at any spatial and temporal scale instantiate the dynamics of adaptive free energy minimization". Whereas the fictional psychohistory merely predicted the behaviour of society, VNE instead begins at the very basis of life itself – cellular life – to argue that it is the self-organizing minimization of information-theoretic free energy which interlinks every conceivable evolutionary and ontogenetic timescale.

To do so, Ramsted et al. give a formal answer to Schrodinger's famous question – 'what is life?' – embedded within the information theoretic framework of the FEP. To live, we are told, is to be an ergodic entity visiting some states with higher probability than others. Further, this ability to resist statistically surprising states depends upon a particular self-organization, which enables an agent to "move, systematically, towards attractive states.... To counter the dispersive effects of random fluctuations". An organism here is nothing more than any set of causal, self-organizing states which

DOI of original article: https://doi.org/10.1016/j.plrev.2017.09.001.
* Correspondence to: Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, United Kingdom.
E-mail address: micah.allen@ucl.ac.uk.

https://doi.org/10.1016/j.plrev.2018.01.007

achieve an autopoietic equilibrium – a Markov blanket of sensory, active, and internal states which both enable the organism to resist entropy and "self-evidence" its own conditions for existence. Bodies and brains are then themselves models of the kinds of embodied actions and environments which are likely to further maximize self-evidence for an organism. Zooming out further still, we find that society, our political institutions, and perhaps even geosphere-level phenomenon such as global warming are simply additional layers in an infinitely nested, Matryoshka-like web of causal links constituting the biological Markov blanket. For Ramstead and colleagues, even evolution is reconfigured under the VNE as a form of meta-Bayesian model selection, acting over the entire nested biological blanket to offer mutual constraints between living entities and their collective behaviours. On this view, the cells, organisms, brains, and societies which make up the world are simply those which successfully maximize the evidence for their existence, rendering the next 'lower' level more predictable. A big claim to be sure – but is it one that can actually be tested empirically?

Answering this question, according to Ramstead and colleagues, requires the genesis of a new interdisciplinary approach to studying the overlap of psychology, biology, and sociology – an FEP-inspired *Foundation*, if you will. This desire to interlink seemingly disparate levels of explanation through the lens of self-organization, though radical, is not without precedent, at least within cognitive science. Indeed, when launching the revolutionary paradigm of enactivism, Francisco Varela wrote:

> "...The enactive program takes a further step... to encompass the temporality of cognition as lived history, whether seen at the level of the individual (ontogeny), the species (evolution), or social patterns (culture) [3]."

For Varela and other enactivists, any attempt to model or understand the human mind which neglects this densely interconnected web of mutual constrainment is doomed to fail. And, much like the VNE, Varellian enactivism emphasizes the cornerstone role of self-organizing, autopoietic dynamics as the crucial linkage across explanatory levels.

Today however, enactivism has largely failed to generalize to mainstream neuroscience and biology, exerting a primarily inspirational influence in the cognitive sciences.[1] Whereas Varela himself saw the approach as building a "methodological remedy for the hard problem [of consciousness]", the multiscale, interdisciplinary zeal of the framework has thus far led to no major empirical revelations and is still relatively niche within the various disciplines of cognitive science and biology. And yet, it is difficult to deny that such a grand, general theory of biological life and intelligence would be of enormous value both to our basic understanding of the world and to society itself.

What underlies this limitation? A key weakness of enactivism is that the framework thus far suffers a general lack of mechanistic falsifiability and empirical prediction. Although enactivists gestured towards unspecified implementations in dynamical systems theory and ecological psychology, cognitive (neuro)scientists seeking to apply these ideas to the brain and behaviour have thus far found little of pragmatic value. That is to say, enactivism offered no clearly falsifiable predictions and argued for no formal mechanisms by which one might realize its ambitions. How exactly one should link between, say high density neural recordings and phenomenological reports, or sociocultural constraints, remains at best vague under the enactive framework. This may be due to the fact that enactivism was by design pitched in opposition to the 'hegemony of the brain', ultimately emphasizing the body and world while having relatively little to say about neurobiology.

Herein lies both the promise and pitfall of Ramstead et al.'s newly forged 'Bayesian enactivism'. The axioms of the FEP, on which the VNE are based, are clearly elucidated mathematical theorems, rather than phenomenological descriptions. The theory does not merely gesture towards dynamical systems, but instead attempts to directly illustrate how specific biological dynamics emerge from a formal information theoretic process. Ramstead and colleagues thus claim that VNE and its application can "be translated into a multidisciplinary research heuristic that promotes a tripartite approach to scientific inquiry". This seems to be at least in principle a valid assumption; as a formal, axiomatic theory, certainly the axioms and proofs on which VNE rests (namely, the free energy principle itself), are subject to analytic refinement, proof, and disproof. But is the theory empirically useful – does it generate concrete, falsifiable hypotheses?

[1] Note that although enactive ideas have generated empirical work, for example in the domain of social interaction [4], cortical dynamics [5], the science of consciousness [6–8], and embodied robotics [9], these have thus far not had much purchase within mainstream neuroscience, biology, or quantitative sociology. Enactivism and other Varelian ideas have no doubt indirectly inspired a generation of scientists and philosophers.