Comment

Beyond blanket terms: Challenges for the explanatory value of variational (neuro-)ethology
Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al.

Jelle Bruineberg [a,b,*], Casper Hesp [b,c,d,1]

[a] *Department of Philosophy, Institute for Logic, Language and Communication (ILLC), University of Amsterdam, Oude Turfmarkt 141, 1012 GC, Amsterdam, Netherlands*
[b] *Amsterdam Brain and Cognition (ABC), University of Amsterdam, Postbus 15900, 1001 NK Amsterdam, Netherlands*
[c] *Institute for Interdisciplinary Studies (IIS), University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands*
[d] *Gravitation Astroparticle Physics Amsterdam (GRAPPA) Institute, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands*

Ramstead et al. [9] integrate the free-energy principle (FEP) [5] and evolutionary systems theory (EST) [1] in order to develop a "meta-theoretical ontology of life", called 'variational neuro-ethology' (VNE). In drawing upon such abstract notions and integrating them even further, they prove themselves to be the ultimate "hedgehogs" [2]: aiming for the ultimate integration of the life sciences and social sciences under one unifying principle. We endorse this pursuit of theoretical integration, especially when derived from first principles. The fundamental nature of their work is exemplified by the book the authors take as their starting point: Schrödinger's *What is Life?* [10]. Given the variety and levels of complexity involved in defining "life", providing an answer to this question is challenging. We first briefly comment on VNE as a label and then highlight some possible problems for the kinds of explanations that would follow from VNE. We address the interrelated charges of (1) merely providing Bayesian and evolutionary "just-so" stories [4,6], and (2) limited interpretative clarity when casting "life" as a series of nested Markov blankets. As a pre-emptive response to these critical remarks, we sketch a few ways forward that we find promising.

A rose by any other name would smell as sweet: The term "neuro-ethology" appears to place unnecessary limits on the framework proposed by Ramstead and colleagues. The prefix "neuro-" refers exclusively to nerve cells, and "ethology" to animal behaviour. This seems to unduly limit the scope of VNE. Given the intended generality of VNE, it applies to non-animal organisms (e.g., plants) *and* to organisms without any nerve cells (e.g., bacteria), as well as to the "behavior" and organization of societies and institutions. The importance of nerve cells for life notwithstanding, neurons are "merely" one type of functional unit in the cross-scale dynamics of life. One of the virtues of FEP is its

applicability to neuronal dynamics *and beyond*. As such, the presentation of this "meta-theoretical ontology of life" could benefit from a label with a more general scope.

VNE aims to provide a rather grand narrative, encompassing all of life. Therefore, concerns may well be raised about explanatory power and practical utility for research. Since VNE is relatively unconstrained and "meta-theoretical", it may be beyond direct empirical refutation, but as a framework it still needs to be able to guide theorizing. Some obstacles emerge here as well. For example, if some levels of organization in life would resist description in terms of "nested Markov blankets", it is unclear whether this would be detrimental to VNE. In the past, both evolutionary theory [6] and Bayesian views of the brain [4] have been accused of providing merely "just-so" stories. Since VNE is a synthesis of both, it faces the same criticism. In our view, VNE could avoid becoming a just-so story if one would demonstrate directly the evolutionary emergence of nested Markov hierarchies, performing approximate Bayesian inference on different temporal and spatial scales. Since the most available (or even only?) way to do so is through elaborate computational modelling, we affirm and appreciate the authors' focus on simulation efforts.

Complex computational models could capture the cross-scale dynamics required to simulate "life" according to VNE. The issue is whether VNE can help interpret such models. When the sheer complexity of models starts to rival that of life itself, we encounter the problem pointed out by J.L. Borges in "Exactitude in Science" [3]. A map so realistic that it is in every way identical to the territory itself would fail to be useful. Ideally, if simulations were to produce an astonishingly complex map of "life", VNE *should* be able to provide useful heuristics to schematize and interpret this map. That utility would provide a measure of the validity of VNE.

VNE conceptualizes life as a series of Markov blankets nested across spatial and temporal scales, where each Markov blanket is a unit of surprisal minimization. The research heuristics suggested by VNE therefore depends crucially both on the identifiability of Markov blankets, and their intelligibility as units of self-organization. The cell boundary of a bacillum is a clear example of a Markov blanket, as are the sensori-motor systems of most organisms. However, one should be careful here: Markov blankets were originally cast by Pearl [8] as epistemological properties (of a specific class) of models rather than as ontological properties of the world. Boundaries as ontological properties that demarcate a spatio-temporal unit of organization are more reminiscent of autopoietic boundaries [7] than of Pearl's original usage.

At any rate, the discretization of *all* of life in terms of Markov blankets appears to be impossible. We consider a few phenomena of life that appear to lack stable statistical insulation between subsystems, implied by the use of the Markov blanket formalism.

In their extension to human culture, Ramstead et al. conceive of a spatio-temporal Markov blanket around the entire species of Homo Sapiens. That is problematic in terms of origins, because a first specimen of Homo Sapiens never existed: we evolved gradually from an ancestral population of primates. In turn, primates evolved gradually from the ancestral population of all mammals. Evolution is continuous and population-wide. It is difficult to define a Markov blanket as enveloping an entire species without including the entire evolutionary tree.

Examples of transient or partial statistical insulation can readily be drawn from human social life. On the one hand, humans certainly do organize in groups which can be statistically insulated from direct influences of other groups (i.e., possess a Markov blanket), for some amount of time. Under VNE, that would count as group "life", with the group performing approximate Bayesian inference concerning the outer world (including other groups). Consider groups of people engaged in war, needing to predict their enemies' next move. On the other hand, human group life can be very transiently organized (e.g., a festival) and with only limited statistical insulation (e.g., internet communities). It is unclear whether these units can be seen as performing an interesting kind of surprisal-minimization and it is hence unclear whether VNE can offer useful research heuristics.

As a closing remark, we are enthusiastic about the integrative potential of VNE and therefore intend these notes as constructive feedback. We have highlighted some of the challenges that, we believe, researchers will encounter when using VNE as a heuristic for future research. The charge of "just-so" stories can be addressed through elaborate computational modelling, but the issues that occur when trying to interpret life as a series of nested Markov blankets require more elaborate formulations and perhaps more stringent criteria than statistical blankets. In particular, VNE will need to address the gaps that arise when Markov blankets are transient or can not be readily identified.

Acknowledgements

References

[1] Badcock PB. Evolutionary systems theory: a unifying meta-theory of psychological science. Rev Gen Psychol 2012;16(1):10.
[2] Berlin I. The hedgehog and the fox. London: Weidenfeld & Nicholson; 1954.
[3] Borges JL. A universal history of infamy. London: Penguin Books; 1946/1975. Translated by Norman Thomas de Giovanni.
[4] Bowers JS, Davis CJ. Bayesian just-so stories in psychology and neuroscience. Psychol Bull 2012;138(3):389.
[5] Friston KJ, Stephan KE. Free-energy and the brain. Synthese 2007;159(3):417–58. https://doi.org/10.1007/s11229-007-9237-y.
[6] Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B, Biol Sci 1979;205(1161):581–98.
[7] Maturana HR, Varela FJ. Autopoiesis and cognition. Berlin: Springer Science & Business Media; 1980.
[8] Pearl J. Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Fransisco: Morgan Kaufmann Publishers; 1988.
[9] Ramstead MJD, Badcock PB, Friston KJ. Answering Schrödinger's question: a free-energy formulation. Phys Life Rev 2018;24:1–16 [in this issue].
[10] Schrödinger E. What is life? Cambridge: Cambridge University Press; 1944.