The free energy principle and autopoiesis: Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead, Paul Benjamin Badcock, Karl John Friston

M. Bitbol[a] and S. Gallagher[b,c]

[a]Archives Husserl, ENS/CNRS, Paris, France
[b]Department of Philosophy, University of Memphis (USA)
[c]Faculty of Law, Humanities and the Arts, University of Wollongong (AU)

Physics of Life Reviews, 24: 24-26, 2018

Ramstead, Badcock and Friston's paper [1] is an ambitious and credible contemporary attempt to answer Schrödinger's question: "What is life?" The authors, however, overemphasize the thermodynamical aspect of the question, while they are content with only general statements about its structural aspect, and tend to ignore the standpoint of the living being itself.

Although much less advanced than the approach taken in the paper by Ramstead et al., Schrödinger's own approach to the process of living was jointly thermodynamical and structural. It raised the question of how living beings can overcome, at least locally, the general tendency to "disorder" implied by the second law of thermodynamics. And it connected this question with some assumptions about the (alleged) flux of "order" coming from the "aperiodic crystal" (now known as macromolecular DNA) that carries genetic information. In Ramstead et al., the thermodynamical status of living beings is studied in exquisite detail, thus convincingly addressing Schrödinger's qualms about the ability of the present laws of physics to make sense of the living. Based on a generalization of the free-energy principle (FEP) and of the "minimization of surprise" condition, formerly used to offer a fruitful theory of neurobiological adaptation, a more general principle meant to apply to every level and time scale of the organization of living beings is proposed. By contrast, the structural aspect of biological processes is only sketched in this paper. It is limited to (i) a hierarchy of levels of organization, and (ii) a general probabilistic characterization (Markov blanket) of the interface between such levels, or between the organism and its environment. By due analogy with the case of theoretical physics, one could say that the theory of the living presented in this paper is a top-down "principle theory" as opposed to a bottom-up "constructive theory" [2]. Like any principle theory, the FEP does not explain *how* a process is accomplished, but

only provides us with a universal model of its development. It overcompensates the opposite tendency of molecular biology, which stays at the level of basic constituents and local structures or processes and overlooks the global organization or global dynamics of living beings.

Yet, the reference to Eigen and Schuster's hypercycles, to self-organizing dynamical systems, and to recursive interactions between components, could have oriented the authors towards a more detailed consideration of the structural patterns that underpin thermodynamics. In particular, it could have prompted consideration of a metabolic phenomenon that is central for maintaining living beings within the "viability set" of their phase space: the self-production of their own constituents out of materials borrowed from their environment; in other terms their *autopoietic* capacities. *Autopoiesis* is a general model proposed by Maturana and Varela that accounts for a wide range of characteristics and behaviors of living beings (including a hierarchy of levels from cell to society), and that also establishes a certain amount of continuity between life and cognition [3, 4]. What kind of relation can we then establish between the two models, namely between the FEP and autopoiesis? To answer that question, we can partly rely on previous work by Friston [5, 6], with some important qualifications.

First, one cannot just say that autopoiesis has been *superseded* by the FEP or that the latter should be 'preferred' over the former, as suggested by Kirchhoff [7]. As we have just suggested, the two models are not exclusive of each other. They differ in epistemological status, and they fulfill distinct theoretical needs. Accordingly they can be used jointly to account for various aspects of the living. For instance, deficiencies in the original autopoietic model can easily be compensated by adding to it some characteristics such as evolutionary adaptation [8], that are also encompassed by the FEP. Furthermore, both models have the capacity to include cognition in the description of the living, each in its own terms but with similar conclusions. Cognition is characterized as the dual approach that a living being must have in order to deal with its environment : (i) integrating inputs from the environment while remaining structurally undisturbed, and (ii) modifying its structure in view of performing such integration. In the augmented autopoietic model this is tantamount to (i) assimilation of molecules within a previously existing circular metabolic network, and (ii) adaptative alterations of the organization of this network [9]. In the FEP, the two steps take

on the form of (i) minimizing surprise and (ii) a Darwinian combination of variation and selection of adaptive processes in view of minimizing surprise [10].

Second, different epistemological presuppositions inform the two models, respectively. For example, autopoiesis, together with its higher scale enactive model, is natively averse to representationalism. In its framework, the environment is not considered as a pre-structured domain that has to be represented by the living unit for the sake of adapting to it. Instead, the environment with its structure and salient features *co-arises* with the process of sense-making of the living unit. Although the FEP is not necessarily antagonistic to this way of thinking about the relation between living beings and their environment [6, 11, 12], the language used by its proponents is often permeated with representationalism [e.g., 5], with a corrective that comes too late to attenuate it. Thus, in section 1.2, the authors evoke the encoding of *external* states by *internal* states, thus presupposing a form of spatial duality. And they speak of organisms that "do not just encode a model of the world, (but *are)* a model of the world", assuming a world out there, independent of the organism, waiting to be modeled. Such a strong representationalist tendency is qualified only in a later sentence: "(an organism is) a physical transcription of causal regularities in its eco-niche that has been sculpted by *reciprocal interactions* between self-organisation and selection over time". The need for this additional qualification strongly contrasts with the autopoietic/enactive model, which was elaborated from the beginning as a systematic alternative to representationalism.

Third, in contrast with the autopoietic/enactive paradigm, it is hard to see in the FEP any affinity with the "existential-phenomenological interpretation of the life-mind continuity thesis" [7, p. 19]. No reference to experience (beyond reference to a neural underpinning) is made, unlike such references in the canonic expositions of autopoiesis [13] and enaction [14]. Moreover, the *compatibility* of the FEP with metabolic selection of salient features of the environment is not sufficient to claim that it natively involves a process of meaning ascription that would allow for the possibility of a first-person standpoint that fits with the third-person model of an organism [15]. As a consequence, unlike the neurophenomenological development of autopoiesis and enaction, the application of the FEP to brain and cognition lacks resources to address issues pertaining to first-person consciousness.

Acknowledgements

SG's research is supported by the Humboldt Foundation's Anneliese Maier Research Award.

References

[1] Ramstead MJD, Badcock PB, Friston KJ. Answering Schrödinger's question: A free-energy formulation. Phys Life Rev (2017), https://doi.org/10.1016/j.plrev.2017.09.001 [this issue]
[2] Ryckman T. Einstein. London: Routledge; 2011
[3] Maturana HR. Varela FJ. Autopoietic systems. *Report BCL* 1975; 9(4): 37-48.
[4] Maturana HR. Varela FJ. Problems in the neurophysiology of cognition. In Autopoiesis and cognition. Dordrecht: Springer; 1980: 41-47.
[5] Friston K. Life as we know it. journal of the Royal Society (London), Interface 2013; 10: 20130475 ; http://dx.doi.org/10.1098/rsif.2013.0475
[6] Allen M. Friston K. From cognitivism to autopoiesis: towards a computational framework for the embodied mind. Synthese 2016; DOI 10.1007/s11229-016-1288-5
[7] Kirchhoff M. Autopoiesis, free energy, and the life-mind continuity thesis. Synthese 2016; https://doi.org/10.1007/s11229-016-1100-6
[8] Di Paolo E. Extended life. Topoi 2009; 28: 9-21.
[9] Bitbol M. Luisi PL. Autopoiesis with or without cognition: defining life at its edge. Journal of the Royal Society (London), Interface 2004; 1: 99-107.
[10] Friston KJ. The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 2010; 11(2): 127-138.
[11] Gallagher S. Allen, M. Active inference, enactivism and the hermeneutics of social cognition. Synthese 2016: 1-22. doi:10.1007/s11229-016-1269-8
[12] Bruineberg J. Kiverstein J. Rietveld E. The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. Synthese 2016: 1-28.
[13] Maturana HR. The organization of the living: A theory of the living organization. Cybernetics Forum 1980; 10 (2-

3): 14-23
[14] Varela F. Thompson E. Rosch E. The embodied mind. Cambridge, MA: MIT Press; 1991.
[15] Vörös S. Bitbol M. Enacting enaction: A dialectic between knowing and being. Constructivist Foundations 2017; 13 (1): 201-210