Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 41 (2022) 61–63

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Comment

A test run of the free energy principle: All for naught? Comment on "How particular is the physics of the free energy principle?" by Miguel Aguilera et al.

Ezequiel A. Di Paolo [a,b,c,*]

[a] *Ikerbasque, Basque Foundation for Science, Bizkaia, Spain*
[b] *IAS-Research Center for Life, Mind and Society, University of the Basque Country, Donostia, Spain*
[c] *Department of Informatics, University of Sussex, Brighton, UK*

Received 5 May 2022; accepted 7 May 2022
Available online 10 May 2022
Communicated by Susan Li

Aguilera et al. [1] have performed an important task in checking the Free Energy Principle (FEP) for coherence and scope. They apply its formalism to a class of tractable models to examine ambiguities and specify when the assumptions and conditions of the FEP warrant the proposed conclusions. The use of idealized models to test theories is standard practice in many areas of science from theoretical physics to macroeconomics (see [2,3] on idealized models in science; see [4–6] for models applied to theories of life and adaptive behavior). One wonders why we had to wait for more than a decade for someone to attempt this kind of "debugging" of the FEP.

For weakly-coupled linear stochastic systems, the FEP conditions are met only in rather special circumstances. The authors suggest these circumstances are rare in living and cognitive systems. I wish to focus on these suggestions because the reasons they give might still be judged as a matter of interpretation by some readers. Let's consider why Aguilera et al. are justified in making them.

Granting convergence to a non-equilibrium steady state (NESS), the FEP assumes a Markov blanket and a lack of solenoidal cross-couplings between internal and external variables (that could take the system out of the steady state). Both conditions put severe limitations on FEP-viable sensorimotor architectures. The canonical configuration connecting the four classes of variables (y, s, a, x; Figure 4A in [1]) most often used in the FEP literature, does *not* meet the necessary requirements. Only what the authors describe as a symmetric "loop" (Figure 4C) works for a second order approximation (higher-order terms could still lead to cross-couplings). Notice that the meaning of "sensory" and "active" states is entirely lost in this peculiar architecture as these states act as a double "barrier" between internal and external states but cannot be easily interpreted in terms of sensors or effectors (for instance, "active" states have no direct coupling with the environment).

Can non-traditional interpretations recover the canonical loop (e.g., by dividing s into new sensor and effectors subclasses)? This will likely lead to a loss of symmetry in the coupling between s and y (J_{sy} and J_{ys}) except if the parameters are just right. But even if this problem could be circumvented the idea requires justifying the plausibility of a "double barrier" between internal and external states. Bacteria use cilia and flagella to feel for obstacles as

DOI of original article: https://doi.org/10.1016/j.plrev.2021.11.001.
* Correspondence to: Ikerbasque, Basque Foundation for Science, Bizkaia, Spain.
E-mail address: ezequiel.dipaolo@ehu.es.

https://doi.org/10.1016/j.plrev.2022.05.001

well as to move them [7], so sensory and active states can indeed be entwined. But the architecture studied by the authors remains at odds with well-known biological processes (horizontal gene-transfer in bacteria, viral infections, mechanical impacts and the effects of gravity, temperature equilibration in cold blooded animals, etc., see [8,9]). These are problems for Markov blankets in general, so can we expect biological systems to be likely to implement an even "stricter" double barrier?

In cognitive systems solenoidal decoupling between internal and external states contradicts empirical findings on non-decomposable systems, e.g., correlations across scales in interaction-dominant dynamics [10], interpersonal synergies [11] and inter-brain synchrony during social interaction [12]. Going into and out of such couplings entails contingent functional and dynamical re-assemblies in brain, body, and world. These phenomena are non-linear, giving weight to Aguilera et al.'s suspicion that the problems they uncover with their linear systems are not likely to improve with more sophisticated models.

When the FEP does work, the authors find it uninformative and unable to capture the history-dependent behavior of the system. Others have also noticed a tension between FEP and historicity [8]. Here it is important to distinguish two senses of history. A strong sense on the one hand, expressed as time translation asymmetry or *path*-dependence [13], e.g., when parameters, constraints, or changing couplings alter the form of the flow or when whole variable sets themselves change. Such phenomena are ubiquitous, from sensorimotor learning and habit formation, to adaptation to permanent injury and metamorphic development. On the other hand lies a weaker sense of history as *time*-dependence. The stronger sense obviously entails the weaker one. The NESS condition is incompatible with the stronger sense of historicity [8] but what about the weaker sense?

Aguilera et al.'s systems obviously evolve over time, but the punchline of the FEP, i.e., the claim that *the* system behaves as if making inferences that track external states by following a gradient that minimizes free energy, doesn't seem to tell us much about the system's actual behavior. What exactly follows this gradient? If the claim applies to the most likely states, this does not hold in general for stochastic systems according to the authors. If what follows the gradient is the conditional average over "counterfactual trajectories", i.e., an *ensemble* average of many runs, we have a problem. This average by definition dilutes the relation between a particular trajectory and the history of system-environment interactions. Hence, even if we restrict ourselves to systems that are not strongly historical, the specific relations of time-dependence (what will *this* system do after X happens?) are also lost in the main claim of the FEP.

Where does the FEP go from here? New versions are likely to appear, hopefully addressing the problems disclosed by Aguilera et al. The proper scope of the FEP may finally be known with some precision, whether organisms in general or some of them in special circumstances as it currently seems. The major lesson, in my opinion, concerns the indispensability of using models as part of the practice of building theories about the big questions of life and mind. Intuitions often fail with complex systems and idealized models are one way to keep them in check, engaging in a sort of "mental calisthenics" [4] to cultivate a feeling for when an idea works. With FEP-related publications numbering in the hundreds, one gets the impression that much effort would have been saved if this exercise had been done earlier.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Thanks to Miguel Aguilera for illuminating conversations on these ideas.

References

[1] Aguilera M, Millidge B, Tschantz A, Buckley CL. How particular is the physics of the free energy principle? Phys Life Rev 2022;40:24–50. https://doi.org/10.1016/j.plrev.2021.11.001.

[2] Morrison M, Morgan M. Models as mediating instruments. In: Morgan M, Morrison M, editors. Models as mediators: perspectives on natural and social science. Cambridge: Cambridge University Press; 1999. p. 10–37.

[3] Reutlinger A, Hangleiter D, Hartmann S. Understanding (with) toy models. Br J Philos Sci 2018;69(4):1069–99. https://doi.org/10.1093/bjps/axx005.

[4] Beer RD. The dynamics of active categorical perception in an evolved model agent. Adapt Behav 2003;2003(11):209–43. https://doi.org/10.1177/1059712303114001.

[5] Beer RD. An investigation into the origin of autopoiesis. Artif Life 2020;26:5–22. https://doi.org/10.1162/artl_a_00307.
[6] Harvey I, Di Paolo EA, Tuci E, Wood R, Quinn M. Evolutionary robotics: a new scientific tool for studying cognition. Artif Life 2005;11:79–98. https://doi.org/10.1162/1064546053278991.
[7] Bloodgood RA. Sensory reception is an attribute of both primary cilia and motile cilia. J Cell Sci 2010;123(4):505–9. https://doi.org/10.1242/jcs.066308. PMID: 20144998.
[8] Di Paolo EA, Thompson E, Beer RD. Laying down a forking path: tensions between enaction and the free energy principle. Philosophy and the Mind Sciences 2022:3. https://doi.org/10.33735/phimisci.2022.9187.
[9] Raja V, Valluri D, Baggs E, Chemero A, Anderson ML. The Markov blanket trick: on the scope of the free energy principle and active inference. Phys Life Rev 2021;39:49–72. https://doi.org/10.1016/j.plrev.2021.09.001.
[10] Anderson ML, Richardson MJ, Chemero A. Eroding the boundaries of cognition: implications of embodiment. Top Cogn Sci 2012;4(4):717–30. https://doi.org/10.1111/j.1756-8765.2012.01211.x.
[11] Riley MA, Richardson MJ, Shockley K, Ramenzoni VC. Interpersonal synergies. Front Psychol 2011;2:38. https://doi.org/10.3389/fpsyg.2011.00038.
[12] Dumas G, Nadel J, Soussignan R, Martinerie J, Garnero L. Inter-brain synchronization during social interaction. PLoS ONE 2010;5(8):e12166. https://doi.org/10.1371/journal.pone.0012166.
[13] Longo G, Montévil M. Perspectives on organisms: biological time, symmetries and singularities. New York: Springer; 2013.