Physics of Life Reviews 47 (2023) 276–278

Contents lists available at ScienceDirect

Physics of Life Reviews

journal homepage: www.elsevier.com/locate/plrev

Comment

From physics to sentience: Deciphering the semantics of the free-energy principle and evaluating its claims
Comment on "Path integrals, particular kinds, and strange things" by Karl Friston et al.

Zahra Sheikhbahaee [a,1], Adam Safron [b,1], Casper Hesp [c], Guillaume Dumas [a,d,*]

[a] CHU Sainte-Justine Research Center, Department of Psychiatry, University of Montreal, Canada
[b] Center for Psychedelic & Consciousness Research, John Hopkins University, United States of America
[c] Department for Developmental Psychology, University of Amsterdam, Netherlands
[d] Mila – Quebec AI Institute, Canada

ARTICLE INFO

Communicated by J. Fontanari

The Free-Energy Principle (FEP) [1–3] has been adopted in a variety of ambitious proposals that aim to characterize all adaptive, sentient, and cognitive systems within a unifying framework. Judging by the amount of attention it has received from the scientific community, the FEP has gained significant traction in these pursuits. The current target article represents an important iteration of this research paradigm in formally describing emergent dynamics rather than merely (quasi-)steady states. This affords more in-depth considerations of the spatio-temporal complexities of cross-scale causality – as we have encouraged and built towards in previous publications (e.g., [4–9]). In this spirit of constructive feedback, we submit a few technical comments on some of the matters that appear to require further attention, in order to improve the clarity, rigour, and applicability of this framework.

Certain idiosyncratic definitions and derivative statements in this paper and prior works have tended to confuse scientists who attempt to assess the validity of such formulations in depth. As the authors aimed to formulate a robust Bayesian mechanics by borrowing from physics, particular care needs to be taken concerning physics terminology.

In the target article, much of the authors' argument relies on what they define as "generalized coordinates", effectively a vector containing higher-order time derivatives of motion of a (sub)system, which can be used to approximate its local dynamics under certain limiting assumptions, on some (unspecified) time scale. While the authors' usage appears to be internally consistent, it may generate confusion with respect to a particular analytical method that physicists hold dear. When grappling with complex dynamical systems, many hours of work can be saved by applying a state-space reduction and exploiting the coordinate-agnostic flexibility of Lagrangian mechanics, where "generalized coordinates" and "generalized velocities" refer to q and $\dot{q}$, respectively, in the Lagrange equation of the second kind: $\frac{d}{dt}(\frac{\delta L}{\delta \dot{q}_j}) = \frac{\delta L}{\delta q_j}$. In that context, carefully choosing a set of generalized coordinates is a highly context-specific and technical task in itself that can simplify derivations of physically meaningful constraints on a system's dynamics (e.g., a textbook example: shifting to angular coordinates to effectively model a pendulum; see [10]). If the authors intend to engage the

DOI of original article: https://doi.org/10.1016/j.plrev.2023.08.016.
* Corresponding author.
E-mail address: guillaume.dumas@umontreal.ca (G. Dumas).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.plrev.2023.11.004
Received 6 November 2023; Accepted 13 November 2023

Available online 17 November 2023
1571-0645/© 2023 Elsevier B.V. All rights reserved.

physics community effectively and avoid unnecessary confusion for novel readers, it would seem expedient to choose a different label, or else clearly explain differences from common (physics-textbook) linguistic use.

Such an apparently innocuous mixing of terms can engender further misunderstandings, leading to breakdowns in communication between fields. For example, consider the following quote from the target article: *"In this setting, the Lagrangian plays the role of an action, where paths of least action minimize the Lagrangian of generalized states." (p.36)* This quote would make most physicists frown because the action is commonly defined as the path integral of the Lagrangian – as acknowledged by the authors (see also [11]). Within the confines of the target article, however, "generalised states" are defined such that they implicitly characterize local paths, which means that the Lagrangian of all derivatives can be seen as specifying an (implicit) action of sorts (valid only to the extent that a Taylor series approximation holds).

More generally speaking, many claims throughout the article contain tacit assumptions, often leaving the reader to reverse engineer what additional details the authors might have had in mind that could help constrain their claims. For example, consider this sentence: *"We make no assumptions about the solutions of this equation, other than the flow operator does not change over some relevant time interval."*(p.38). Since the formulation itself consists of postulates and approximations with unclear boundary conditions, it becomes difficult (or impossible) to evaluate individual claims. In their pursuit of generality, most concepts have been left either unconstrained or defined in a self-referential manner. In extension, the same holds for solutions, as exemplified above: the authors assumed that solutions have flow operators that do not change on relevant time intervals without specifying what degrees of change might be considered significant and what kinds of time intervals could be "relevant". One might wonder under which conditions a Taylor series will (still) provide satisfactory approximations when considering systems that display cross-scale itinerancy. There are likely dynamical regimes where such polynomial approximations must be extended with periodic components, but currently, the blanket response would be to simply postulate "relevant" time scales at which the Taylor series is enough.

The need for precise semantics becomes critical when the FEP claims delve into complicated topics such as agency and sentience. The words autonomy and agency are used interchangeably, but there is a significant difference between systems driven by meaningful internal dynamics [12], and systems that are capable of planning in order to achieve explicitly represented goals; thus, it remains important to recognize the difference between teleology vs. 'mere' telonomy [13–15].

In the article, the authors invoke notions of "strange particles" and the circular causation involved in their beliefs about their self-generated actions. This handling of "self-evidencing" [16] may not only help speak to the "strange loops" characteristic of some complex adaptive systems, but also to the deeper issues of life-mind continuity [17], and the recursive nature of consciousness [18,19].

The authors also allude to the concept of consciousness (using quotes from Heisenberg and David Bohm), but never clarify how it relates to their definition of "sentience". Connecting the behaviour of strange particles with 'sentience' might be premature, depending on the semantic scope of this concept. While self-evidencing systems can be considered to be sentient in that they are governed by information given valence by reference to self-persistence, it is important not to conflate this with consciousness, lest we risk being so inclusive with this concept as to compromise the utility of considering the conscious properties of a system. If conscious brains are clearly a self-referential system, not all self-referential systems are necessarily sentient or conscious.

Conscious experience likely requires complex network architectures capable of supporting "integrated world modelling" [20,21]. Such functional properties may potentially be achieved in a variety of ways in different kinds of "particles," such as through information bottlenecks emerging from complex attractor dynamics [22]. However, to reliably ground such inferences about the potential conscious properties of a system, it is required that we both deeply engage with models of consciousness for systems for which we know there are subjective experiences, as well as precisely characterize the phenomenology (in the sense used by physicists) of systems that may realize such qualitative aspects by other means."

To associate the functionalities (and perhaps phenomenality) of "sentience" with particles may generate further confusion by both implying the kinds of panpsychism the authors explicitly renounce, as well as potentially suggesting a kind of illegitimate eliminative reductionism concerning the values of agents with conscious experience (e.g. the "as if" language used throughout the paper for the ontological status of particles with goals).

In conclusion, this FEP-based formalism presented by Friston and colleagues offers an intriguing perspective on life and cognition, but many conceptual nuances still need to be addressed. This painstaking process could eventually allow the authors to do justice to the complex systems they want to describe and the global and interdisciplinary community of academics they aim to engage.

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.

Acknowledgement

Zahra Sheikhbahaee and Adam Safron were supported by the Survival and Flourishing Fund (SFF-2023-H1) through the Institute for Advances Consciousness Studies (IACS). Guillaume Dumas was supported by the Institute for Data Valorization, Montreal (IVADO; CF00137433), the Fonds de recherche du Québec (FRQ; 285289), the Natural Sciences and Engineering Research Council of Canada (NSERC; DGECR-2023-00089), and the Azrieli Global Scholars Fellowship from the Canadian Institute for Advanced Research (CIFAR) in the Brain, Mind, & Consciousness program. Additionally, the authors gratefully acknowledge the helpful feedback provided by colleagues, particularly Vincent Taschereau-Dumouchel. The author has no conflicts of interest to disclose.

References

[1] Friston K, Costa LD, Sakthivadivel D, Heins C, Pavliotis G, Ramstead M, et al. Path integrals, particular kinds, and strange things. Phys Life Rev 2023;47:35–62.
[2] Friston KJ, Stephan KE. Free-energy and the brain. Synthese 2007;159:417–58. https://doi.org/10.1007/s11229-007-9237-y.
[3] Friston K. A free energy principle for a particular physics. arXiv:1906.10184, 2019.
[4] Bruineberg J, Hesp C. Beyond blanket terms: challenges for the explanatory value of variational (neuro-)ethology: Comment on "Answering Schrödinger's question: a free-energy formulation". Phys Life Rev 2018;24:37–9. https://doi.org/10.1016/j.plrev.2017.11.015.
[5] Hesp C, Hipólito I. Living on the edge – practical information geometry for studying the emergence and propagation of life forms: Comment on "How particular is the physics of the free-energy principle?" by Aguilera et al. Phys Life Rev 2022;42:52–5. https://doi.org/10.1016/j.plrev.2022.06.004.
[6] Hesp C. Spatiotemporal constraints of causality: blanket closure emerges from localized interactions between temporally separable subsystems. Behav Brain Sci 2022;45:e197. https://doi.org/10.1017/S0140525X22000620.
[7] Friston K, Costa LD, Hafner D, Hesp C, Parr T. Sophisticated inference. Neural Comput 2021;33(3):713–63. https://doi.org/10.1162/neco_a_01351.
[8] Kastel N, Hesp C. Ideas worth spreading: a free energy proposal for cumulative cultural dynamics. In: Machine learning and principles and practice of knowledge discovery in databases. Springer International Publishing; 2021. p. 784–98.
[9] Kelso J, Dumas G, Tognoli E. Outline of a general theory of behavior and brain coordination. Neural Netw 2013;37:120–31. https://doi.org/10.1016/j.neunet.2012.09.003.
[10] Goldstein H, Poole C, Safko J. Classical mechanics. Addison-Wesley; 2002.
[11] Landau L, Lifshitz E. Mechanics. Elsevier; 1976.
[12] Rovelli C. Meaning = information + evolution. arXiv:1611.02420, 2016.
[13] Humphrey N. Sentience: the invention of consciousness. MIT Press; 2023.
[14] Safron A. The radically embodied conscious cybernetic Bayesian brain: from free energy to free will and back again. Entropy 2021;23(6). https://doi.org/10.3390/e23060783.
[15] Deacon TW, Holsopple B. Incomplete nature: how mind emerged from matter. W. W. Norton & Company; 2011.
[16] Hohwy J. The self-evidencing brain. Noûs 2016;50(2):259–85. https://doi.org/10.1111/nous.12062.
[17] Thompson E. Mind in life: biology, phenomenology, and the sciences of mind. Harvard University Press; 2010.
[18] Hofstadter Gödel, Escher D. Bach: an eternal golden braid. Basic Books; 1979.
[19] Thompson E, Varela F. Radical embodiment: neural dynamics and consciousness. Trends Cogn Sci 2001;5(10):418–25. https://doi.org/10.1016/S1364-6613(00)01750-2.
[20] Safron A. Integrated world modeling theory expanded: implications for the future of consciousness. Front Comput Neurosci 2022;16:642397. https://doi.org/10.3389/fncom.2022.642397.
[21] Safron A. An integrated world modeling theory (IWMT) of consciousness. Front Artif Intell 2020;3:30. https://doi.org/10.3389/frai.2020.00030.
[22] Volzhenin K, Changeux J, Dumas G. Multilevel development of cognitive abilities in an artificial neural network. Proc Natl Acad Sci 2022;119(39):e2201304119. https://doi.org/10.1073/pnas.2201304119.