Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews ••• (••••) •••–•••

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Comment

Some minimal notes on notation and minima A comment on "How particular is the physics of the free energy principle?" by Aguilera, Millidge, Tschantz, and Buckley

Maxwell J.D. Ramstead [a,b], Dalton A.R. Sakthivadivel [c,b]

[a] *Wellcome Centre for Human Neuroimaging, University College London, London WC1N 3AR, UK*
[b] *VERSES Research Lab and Spatial Web Foundation, Los Angeles, CA 90016, USA*
[c] *Department of Mathematics, Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794-3651, USA*

Received 5 May 2022; accepted 11 May 2022

Communicated by Susan Li

Abstract

We comment on a technical critique of the free energy principle in linear systems by Aguilera, Millidge, Tschantz, and Buckley, entitled "How Particular is the Physics of the Free Energy Principle?" Aguilera and colleagues identify an ambiguity in the flow of the mode of a system, and we discuss the context for this ambiguity in earlier papers, and their proposal of a more adequate interpretation of these equations. Following that, we discuss a misinterpretation in their treatment of surprisal and variational free energy, especially with respect to their gradients and their minima. In sum, we argue that the results in the target paper are accurate and stand up to rigorous scrutiny; we also highlight that they, nonetheless, do not undermine the FEP.

Keywords: Free-energy principle; Active inference; Particular physics

1. Preliminary remarks

We are delighted to comment on the interesting technical critique of the free energy principle (FEP) by Aguilera, Millidge, Tschantz, and Buckley. This critique is noteworthy for several reasons. Amongst them are the technical rigour and astuteness of the work done, for which [1] is distinct in the literature. We are especially pleased to see that the paper has provoked a dialogue which is both friendly and productive. Indeed, although we disagree with some aspects of the critique, we contend that, in the years since the paper was originally circulated as a preprint, our collective understanding of the core formalism and scope of the FEP has increased significantly—in no small part

DOI of original article: https://doi.org/10.1016/j.plrev.2021.11.001.
E-mail addresses: maxwell.ramstead@verses.io (M.J.D. Ramstead), dalton.sakthivadivel@stonybrook.edu (D.A.R. Sakthivadivel).
URL: https://darsakthi.github.io (D.A.R. Sakthivadivel).

https://doi.org/10.1016/j.plrev.2022.05.005
1571-0645/

due to the conversations fostered by this critique. The community as a whole has undoubtedly benefited from this exchange, and we are grateful to have been part of it.

In [1], the authors present a detailed and rigorous analysis of the free energy principle. In this comment, we focus on two of the points made in [1]: one regarding marginal flows and their average, the other, regarding surprisal and variational free energy gradients, and their minima. We argue that, whilst the results in the target paper are accurate and stand up to rigorous scrutiny, they do not undermine the FEP. Lastly, we note that this was clearly by design: the work in the target paper was never meant to undermine the FEP. Rather, it was meant to test its applicability to—and informativeness about—linear systems.

2. Flows of averages and averages of flows

Many of the core papers in the FEP literature (seem to) equate the average flow of a system to the flow of the average. It is certainly true that the average of the flow does not equal the flow of the average in general; nor even is this true generically. The authors are absolutely correct when they make this point. It is likely this error can be traced back to an unfortunate notational choice in [4], where equation 3.5 therein reads

$$\dot{\boldsymbol{\eta}}(b) = (Q - \Gamma)\nabla\mathfrak{I}(\boldsymbol{\eta}, b).$$

This equation ought to have been written

$$\langle\dot{\eta}(b_t)\rangle = (Q - \Gamma)\nabla\mathfrak{I}(\boldsymbol{\eta}_t, b_t), \tag{1}$$

with $\langle\cdots\rangle$ denoting an expectation. The given decomposition of the flow is the deterministic component of a stochastic differential equation, arising from an averaging over the random fluctuations in the SDE perturbing its flow. That is, our equation (1) expresses (in clearer notation) the observation that the flow of the expected external state is predicated on the current blanket state and expected external state at a given time-point, and drifts based on the gradient of the surprisal of those variables, as calculated at that time point.

That is precisely what is claimed in the FEP: that the flow of the mode is on average the drift component of an SDE for that flow. What is labelled as Assumption 3* in the target paper is not assumed by the FEP. Our equation (1), which denotes equation 3.5 [4] more properly, makes it apparent that one cannot read the flow operator as commuting with the expectation operator, which undermines Assumption 3* as a valid interpretation of the FEP.

This should not, however, be construed as a mistake on the part of the authors. On the contrary, the authors reproduce the intended definition of marginal flow in [4], which they label Assumption 3**. The point of critique raised here is only that their initial claim about the flow of the average and average of the flow in Assumption 3* is a *non sequitur*.

To summarise, the results of the target paper show that it would be nonsensical to interpret the FEP as asserting Assumption 3*, as it would lead to incorrect results. Luckily, the FEP does not depend on this assumption—despite, perhaps, being misleadingly written in some places. This is a purpose the critique serves decidedly well. It demonstrates rather clearly some pitfalls in the FEP literature, especially in the way it has been written in some core papers, and thus, the way it must not be read.

Finally, since their analysis is predicated on Assumption 3** as though in retrospect, other results are independent of these statements.

3. Gradients of surprisal and gradients of free energy

The mishandling the definition of marginal flows in [4] touches indirectly on a second issue in the claims of the target paper, which we do take to be mistaken—namely, that the gradient of free energy ought to be informative about the average flow of external states, present in equation A9 of [1].

It is obvious that these gradients exist in different spaces—hence the dual information geometry spoken of in [3]—but to the authors' credit, their argument is more subtle: the two gradients are linearly related when σ is a linear function. However, here we note that linear transformations do not in general preserve the shape of a vector field, nor the shape of flows in that field. They simply map an identity element to an identity element. In particular, it is claimed in the target paper that when σ acts linearly on the gradient of surprisal, it maps this vector field to the gradient of free energy linearly, and so we should be able to understand external states as flowing along the gradient of free energy

just as well as along the gradient of surprisal. In fact, it is only the case that these vector fields share a minimum given by σ, and not that they reach those minima at the same time, nor for the same states.

One can imagine a relatively simple counterexample to the foregoing statement that the gradient flow of one quantity is meaningful to a quantity which exhibits a linearly related gradient flow. Take, for instance, the planar vector fields $X = (-x, -y)$ and $Y = (y, -x)$, and the following linear transformation T relating them:

$$\begin{bmatrix} \cos\frac{1}{2}\pi & -\sin\frac{1}{2}\pi \\ \sin\frac{1}{2}\pi & \cos\frac{1}{2}\pi \end{bmatrix} \begin{bmatrix} -x \\ -y \end{bmatrix} = \begin{bmatrix} y \\ -x \end{bmatrix}.$$

We certainly have $X_{(0,0)} = (0, 0)$ and $Y_{(0,0)} = (0, 0)$, and indeed, $T(0, 0) = (0, 0)$. Hence, the zero point of Y is the zero point of Y. However, the two gradients are very different, and their integral curves—the flows in that gradient—will also be quite different. This is not even to mention that when T is non-linear—even an affine map, a linear map with a constant shift, will do—their minima will often occur at different states.

In point of fact, exactly one such counterexample appears in the target paper: it is demonstrated that the gradient of $F(\boldsymbol{\eta}, b)$ does not resemble the gradient of $\mathfrak{I}(\boldsymbol{\eta}, b)$. Once again the authors are absolutely correct, but do not adequately connect their objection to the FEP. It is true, as they claim, that equating the two cannot be done; but nowhere does the FEP claim it can be done, nor does the FEP expect to draw insights from doing so.

What the line of reasoning that the authors took *would* allow us to say is that $\boldsymbol{\eta}$ flows on $\nabla_\eta \mathfrak{I}(\boldsymbol{\eta}, b)$ whilst the *linear transformation* of the flow of $\boldsymbol{\eta}$, here $\nabla\sigma^{-1}(\boldsymbol{\eta})$, flows on $\nabla_\mu \mathfrak{I}(\sigma(\boldsymbol{\mu}), b)$. This is equivalent to the existence of a relation

$$(\nabla_\mu \sigma)^{-1} \nabla_\mu \mathfrak{I}(\sigma(\boldsymbol{\mu}), b) = \nabla_\eta \mathfrak{I}(\boldsymbol{\eta}, b) \tag{2}$$

between the two flows, which can indeed be produced by a simple application of the chain rule and $\sigma(\boldsymbol{\mu}) = \boldsymbol{\eta}$. Equation (2) tells us that the flows in these vector fields are linearly related, but are emphatically *not* identical. Respecting (2), the relation between flows holds in the same general sense as $T(x, y) = (x', y')$ for arbitrary pairs of vectors—not just zero points—but, it is concomitant on not equating one to the other at any point other than the zero point. Correspondingly, we note that $\boldsymbol{\eta}$ under the linear transformation σ^{-1} is $\boldsymbol{\mu}$, and that

$$\nabla_\mu \mathfrak{I}(\sigma(\boldsymbol{\mu}), b) = \nabla_\mu F(\boldsymbol{\mu}, b),$$

recovering the equations in [4] *and* illuminating the claim that

$$\nabla_\mu F(\boldsymbol{\mu}, b) \neq \nabla_\eta \mathfrak{I}(\boldsymbol{\eta}, b)$$

in [1]. The proper application of the synchronisation map σ is known to generate the desired vector fields and flows, and in particular, relates the flow of $\boldsymbol{\mu}$ on the surprisal $\nabla_\mu \mathfrak{I}(\sigma(\boldsymbol{\mu}), b)$ to a flow on the particular free energy $\nabla_\mu F(\boldsymbol{\mu}, b)$, seen in [2].

As such, the FEP says nothing of free energy over external states, and avoids doing so by design. Here we arrive at an important critical remark: more generally, the FEP says nothing about whether systems actually do really minimise free energy gradients, or whether we can merely describe them as such; nor whether systems minimise their free energy, as opposed to merely tending towards a free energy minimum, and thus admitting an equivalent model of minimal free energy.

In closing, we are grateful to have been able to comment on this paper. We believe that the ensuing exchange will be part of a new, productive phase of development for the FEP. Consistent with the aim of the target paper, asking difficult technical questions of the FEP and searching for the answers is always a fruitful scientific exchange.

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.

References

[1] Aguilera Miguel, Millidge Beren, Tschantz Alexander, Buckley Christopher L. How particular is the physics of the free energy principle? Phys Life Rev 2022;40:24–50.

[2] Da Costa Lancelot, Friston Karl J, Heins Conor, Pavliotis Grigorios A. Bayesian mechanics for stationary processes. Proc R Soc A 2021;477(2256):20210518.
[3] Friston Karl J. A free energy principle for a particular physics. Preprint arXiv:1906.10184, 2019.
[4] Parr Thomas, Da Costa Lancelot, Friston Karl J. Markov blankets, information geometry and stochastic thermodynamics. Philos Trans R Soc A 2020;378(2164):20190159.