Please cite as: Veit, W. & Browning, H. (2022). Life, Mind, Agency: Why Markov Blankets Fail the Test of Evolution. *Preprint*.

Check www.walterveit.com for citation details once published

Life, Mind, Agency: Why Markov Blankets Fail the Test of Evolution

Walter Veit & Heather Browning

Abstract:

There has been much criticism of the idea that Friston's free energy principle can unite the life and mind sciences. Here, we argue that perhaps the greatest problem for the totalizing ambitions of its proponents is a failure to recognize the importance of evolutionary dynamics and to provide a convincing adaptive story relating free energy minimization to organismal fitness.

Main text:

In the recent explosion of literature on the free energy principle, many authors have become increasingly frustrated with the grand ambitions toward using it as a general and unified theory of life, mind, and agency. While many have noted the gulf between the mathematical framework of the free energy principle and its application to real target systems, in their target article Bruineberg et al. (2021) offer what is perhaps the most detailed and sustained criticism of the use of Markov blankets in the biological and cognitive sciences. They argue against what they consider an imprecise use in these sciences for defining entities such as organisms, agents, and minds, differentiating between the theoretical 'Pearl blankets' and the more metaphysically-laden 'Friston blankets'. Since these two interpretations are often confused and those making metaphysical claims often retreat to an instrumentalist view once pushed, Bruineberg et al. have provided us with a useful tool to distinguish inferences within the model from inferences with a model, which ought not to be done based on the usefulness of the mathematical framework alone. We welcome the challenge to a perceived conflation between the in-principle applicability of the mathematical framework to any self-organizing system and to the conviction that Markov blankets are able to revolutionize our understanding of the living world (Friston 2013).

The authors note that a realist reading of the application of Friston blankets requires not just the mathematical frameworks established for the use of Pearl blankets, but also independent metaphysical assumptions that, they argue, have not yet been provided. Here, we wish to build on this point by emphasizing the need for these assumptions to align with a plausible Darwinian story. We argue that one of the major problems in recent attempts to use Markov blankets to define the boundaries of

organisms and their environments is that they fail to pass the bottleneck of evolutionary theory and give us a misleading picture of living agents and what they are *for*.

Bruineberg et al. (2021) show that one cannot just 'read off' the boundary between agent and environment from the mathematical formalism provided in the theoretical models. Instead, these are ambiguous and depend on additional assumptions by the modeler, thus requiring quite substantive metaphysical supplementation for Markov blankets to do their work. Here they note that one of the ways of picking out the 'right' model for identifying the ontologically significant Friston blanket is through use of the free energy principle – relying on the assumption that living systems aim at minimizing free energy. It is this basic assumption of the free energy principle that we wish to challenge. This framework fails to demarcate the organismal boundary that *matters*, from an evolutionary point of view.

As philosophers such as Ruth Millikan and Dan Dennett have long argued, it is only by paying attention to the theoretical bottleneck of evolutionary theory that we can distinguish important properties, boundaries, and processes of living systems between those that *matter* to the organism from those that do not. Markov blankets are said to be able to identify the boundaries of any agent in the sense of a self-organizing system (Ramstead et al. 2019), but they fail to distinguish the right boundaries to understand the evolution of living systems. It has been an oversight within Friston's framework to fail to engage with evolutionary theory and the question of what the organism is *for*. It is only in this *teleonomic* context, that we can make sense of the functional boundaries of life, mind, and agency as properties of biological systems. As the framework fails to answer the hard question of why it is the properties picked out by attempts to apply Markov blankets to biological systems, it cannot succeed in both its explanatory and metaphysical ambitions.

The question that this framework would need to answer in order to be successful in this biological context, is what is the adaptive function of minimizing free energy? That is, how does this process contribute to the survival and reproduction of the organism? One response may be to simply assert that adaptive fitness and negative free energy are "the same thing" (Friston et al. 2012, p. 2). However, it is not clear why one should take this to be true – predictive expectations and fitness values do not on their surface appear to constitute anything like the same thing. Another path may be to argue instead that minimization of free energy, while not *constituting* fitness, is still a strong *contributor* to it, in that organisms that act in this way will typically have higher survival and reproduction. However, again, it is not immediately clear why one should believe this. As an example of why this is not particularly plausible, take the *Dark Room Problem*, which offers the challenge that prediction error would be best minimized through sitting still in a dark room, but organisms clearly did not evolve this way (Mumford 1992; Clark 2013). If we treat all of the cognitive activities of organisms as a form of prediction or surprise minimization, there will inevitably be "a wedge between what is typical and what is good" (Klein 2018, p. 2548); we should instead allow that there may be other functions that will not always align with prediction minimization. We then need a more detailed description of the fitness benefits, and how they might be weighted or traded off against other adaptive functions of an organism.

As well as the problems described by the authors of mistaking the useful abstraction Markov blankets provide for the purposes of Bayesian modelling with the idea that free energy minimization is all that goes on in living systems, we add what is perhaps the greatest problem in the biological context: that it forces us to idealize away from the most important features of living organisms and thus will provide

a false and diminished picture of the world. Without the recognition of the importance of evolutionary dynamics, the totalizing ambitions of the free energy principle to unite the mind and life sciences must fail.

Conflict of interest statement:

The authors have no conflicts of interest to report.

Funding Information:

WV's research was supported under Australian Research Council's Discovery Projects funding scheme (project number FL170100160).

References

Bruineberg, J., Dolega, K., Joe Dewhurst, Manuel Baltieri (2021). The Emperor's New Markov Blankets. Behavioral and Brain Sciences.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–253

Friston, K. J. (2013). Life as we know it. Journal of the Royal Society Interface, 10(86): 20130475.

Friston, K., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. Frontiers in Psychology, 3, 1–7.

Klein, C. (2018). What do predictive coders want? Synthese, 195(6), 2541-2557.

Mumford, D. (1992). On the computational architecture of the neocortex. Biological Cybernetics, 66(3), 241–251.

Ramstead, M. J., Kirchhoff, M. D., Constant, A., and Friston, K. J. (2019). Multiscale integration: beyond internalism and externalism. Synthese, 198:41–70.