Ikegami, 2009). A biological boundary, then, is not a given, but is actively defined by the system itself. The same principle applies across a wide variety of living systems, from single cell creatures to more complex, multicellular animals, which live sensorimotor lives (Thompson, 2007).

Cognitive systems likewise actively produce their boundaries through their interaction with the environment. We can see this in the case of extended cognition and mind (Clark, 2008; Clark & Chalmers, 1998), where cognitive boundaries extend beyond the biological body by incorporating environmental items as their constitutive parts. Cognitive extension is not a state upon which we stumble by chance; rather, it is a process we actively bring about, or "enact," based on skills and habits cultivated over time (Miyahara & Robertson, 2021; Miyahara, Ransom, & Gallagher, 2020). To illustrate, consider Otto from Clark and Chalmers' (1998) famous thought experiment. Otto suffers a mild case of Alzheimer's disease and uses a notebook to compensate for his memory deficit. According to Clark (2010), Otto and his notebook exhibit a tight functional coupling with each other to constitute a unified cognitive system (Miyazono, 2017) to the extent that they satisfy the following "trust and glue" conditions: (1) the resource (viz., the notebook) is reliably available and typically invoked; (2) any information thus retrieved is more or less automatically endorsed; and (3) information contained in the resource is easily accessible as and when required. Obviously, Otto will not meet these conditions merely by developing a memory problem. Rather, he would have to learn to use notebooks to complement his compromised cognitive capacities and continue to do so repeatedly until it became a habit for him to always carry around a notebook and use it for constant notetaking. The functional coupling is a product of Otto's active engagement with the notebook and his development of relevant skills and habits over time (which is why Otto's Markov blanket is malleable [Clark, 2017] or negotiable [Kirchhoff and Kiverstein, 2021]).

The main shortcoming of the Friston blanket approach concerns the relationship between action (i.e., active inference) and identity (i.e., the Markov blanket). In this approach, active inference depends upon the Markov blanket, but not the other way round. Biological and cognitive systems are defined by Markov blankets as boundaries with the external environment. These systems perform active inferences through looping interactions between sensory states, internal states, and active states defined by the Markov blanket to keep their internal parameters within viable bounds (Friston, 2013). On the other hand, as we saw above, both biological and cognitive systems actively create and maintain their bounded identity by interacting with the environment. As Clark puts it: "Creatures like us […] are Nature's experts at knitting their own Markov blankets" (Clark, 2017, p. 14). To accommodate this within the free-energy principle (FEP) framework, we must conceive of active inference as playing an essential role in autopoiesis, that is, in creating and maintaining the system's bounded identity (cf. Kirchhoff, 2018). In fact, Friston (2010) describes living systems as performing active inference to reduce sensory surprisal and consequently maintain its homeostasis. Nevertheless, on the FEP, active inference does not explicitly participate in the autopoietic formation of the boundary between the system and its environment, which defines the identity of living beings, that is, the Markov blanket. That is, the dynamic relationship between action and identity is missing in the Friston blanket approach that depicts Markov blankets not as a product, but only as a precondition of active inference (Friston, 2013).

In short, the Friston blanket approach fails to identify the tailor who creates the boundaries. At most, Markov blankets coincide with the outcome of the boundary-making processes carried out by biological and cognitive agents. Markov blankets are tailored by statistical patterns but living agents do not outsource boundary-making: We actively weave our own boundaries with the world.

Acknowledgments. We are grateful to Masatoshi Yoshida for his inputs during the preparation stage of this commentary.

Financial support. This work was supported by the JSPS KAKENHI Grant Number 20K00001.

Conflict of interest. None.

References

Clark, A. (2008). *Supersizing the mind*. Oxford University Press.

Clark, A. (2010). Memento's revenge: The extended mind, extended. In R. Menary (Ed.), *The extended mind* (pp. 43–66). MIT Press.

Clark, A. (2017). How to knit your own Markov blanket: Resisting the second law with metamorphic minds. In T. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing*: 3. MIND Group. https://doi.org/10.15502/9783958573031

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19.

Egbert, M. D., & Di Paolo, E. (2009) Integrating autopoiesis and behavior: An exploration in computational chemo-ethology. *Adaptive Behavior*, *17*(5), 387–401.

Friston, K. (2010) The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138.

Friston, K. (2013). Life as we know it. *Journal of The Royal Society Interface*, *10*(86), 20130475. http://dx.doi.org/10.1098/rsif.2013.0475

Ikegami, T., & Suzuki, K. (2008). From a homeostatic to a homeodynamic self. *BioSystems*, *91*(2), 388–400.

Kirchhoff, M. D. (2018). Autopoiesis, free energy, and the life–mind continuity thesis. *Synthese*, *195*(6), 2519–2540.

Kirchhoff, M. D., & Kiverstein, J. (2021). How to determine the boundaries of the mind: A Markov blanket proposal. *Synthese*, *198*(5), 4791–4810.

Miyahara, K., Ransom, T. G., & Gallagher, S. (2020). What the situation affords: Habit and heedful interrelations in skilled performance. In F. Caruana & I. Testa (Eds.), *Habit: Pragmatist approaches from cognitive neurosciences to social sciences* (pp. 120–136). Cambridge University Press.

Miyahara, K., & Robertson, I. (2021) The pragmatic intelligence of habits. *Topoi*, *40*, 597–608.

Miyazono, K. (2017). Does functionalism entail extended mind?. *Synthese*, *194*(9), 3523–3541.

Suzuki, K., & Ikegami, T. (2009). Shapes and self-movement in protocell systems. *Artificial Life*, *15*(1), 59–70.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Varela, F. R. (1979). *Principles of biological autonomy*. North Holland.

Varela, F. R., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, *5*, 187.

Life, mind, agency: Why Markov blankets fail the test of evolution

Walter Veit[a] and Heather Browning[b]

[a]School of History and Philosophy of Science, The University of Sydney, Sydney, NSW 2006, Australia and [b]London School of Economics and Political Science, Centre for Philosophy of Natural and Social Science, London WC2A 2AE, UK
wrwveit@gmail.com
DrHeatherBrowning@gmail.com
https://walterveit.com/
https://www.heatherbrowning.net/

doi:10.1017/S0140525X22000115, e214

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.

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. 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." As 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. 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, Kirchhoff, Constant, & Friston, 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, Thornton, & Clark, 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 (Clark, 2013; Mumford, 1992). 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 modeling 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.

Financial support. W. V.'s research was supported under Australian Research Council's Discovery Projects funding scheme (project number FL170100160).

Conflict of interest. None.

References

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

https://doi.org/10.1017/S0140525X21002351 Published online by Cambridge University Press

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., & Friston, K. J. (2019). Multiscale integration: Beyond internalism and externalism. *Synthese*, *198*, 41–70.

Embracing sensorimotor history: Time-synchronous and time-unrolled Markov blankets in the free-energy principle

Nathaniel Virgo[a], Fernando E. Rosas[b,c,d,e] and Martin Biehl[f]

[a]Earth-Life Science Institute (ELSI), Tokyo Institute of Technology, Tokyo 152-8550, Japan; [b]Department of Brain Science, Centre for Psychedelic Research, Imperial College London, London W12 0NN, UK; [c]Data Science Institute, Imperial College London, London W7 2AZ, UK; [d]Centre for Complexity Science, Imperial College London, London W7 2AZ, UK; [e]Department of Informatics, University of Sussex, Brighton BN1 9QJ, UK and [f]Cross Labs, Cross Compass, Tokyo 104-0045, Japan
nathanielvirgo@elsi.jp
f.rosas@imperial.ac.uk
martin.biehl@cross-compass.com
http://www.elsi.jp/en/members/researchers/nvirgo
https://www.imperial.ac.uk/people/f.rosas
https://twitter.com/36zimmer

doi:10.1017/S0140525X22000334, e215

Abstract

The free-energy principle (FEP) builds on an assumption that sensor–motor loops exhibit Markov blankets in stationary state. We argue that there is rarely reason to assume a system's internal and external states are conditionally independent given the sensorimotor states, and often reason to assume otherwise. However, under mild assumptions internal and external states are conditionally independent given the sensorimotor *history*.

Bruineberg and colleagues provide a thorough review of Markov blankets and their limitations in the context of the free-energy principle (FEP). We wish to complement this by drawing attention to two additional issues that we believe have important consequences for the FEP.

Firstly, contrary to what one might expect, the condition known as "Markov blanket" in the FEP literature is generally not guaranteed by a sensor–motor loop structure. Secondly, the Markov blanket condition needed for the FEP is far stronger than it appears to be. These issues severely limit the scope of applicability of current formulations of the FEP. Fortunately, we believe they can be solved, and give some hints towards a resolution.

As Bruineberg et al. explain, the notion of a Markov blanket arises in the context of graphical models, and in particular, Bayesian networks. In a Bayesian network each node represents a random variable, and their joint distribution factors in a particular way that depends on the topology of the graph (Pearl, 1988).

The literature on FEP is also concerned with graphs that are not Bayesian networks. Each node in these graphs represents a dynamical variable of a system and an edge represents the possibility that one dynamical variable can influence another. These include the adjacency matrix described in Bruineberg et al.'s section 4.2, and also the sensor–motor loop as illustrated in their Figure 2. Typically, the edges in such graphs correspond to non-zero terms in a Jacobian matrix. We will call such graphs *influence graphs*.

A stationary state defines a joint distribution over the nodes of an influence graph. There is then some resemblance between the influence graphs and Bayesian networks, since both contain nodes that represent random variables and edges that represent influences of some kind.

However, these two types of graph are fundamentally different. Influence graphs are not necessarily acyclic, but more importantly, the theorems in Pearl's formalism do not apply to influence graphs. In particular, one might expect that the sensor–motor loop (Bruineberg et al.'s Fig. 2) would imply the *time-synchronous Markov blanket condition*

$$\mu_t \perp\!\!\!\perp \phi_t \mid s_t,\ a_t. \tag{1}$$

However, this is not the case in general – and this is important because (1) is used in deriving the FEP. This issue has been recently pointed out (Aguilera, Millidge, Tschantz, & Buckley, 2021; Biehl, Pollock, & Kanai, 2021), and while it has been acknowledged in some of the most recent FEP literature it is not as widely known as it should be. We sketch the underlying reason for it in Figure 1.

Recent works (e.g., Friston, Heins, Ueltzhöffer, Da Costa, and Parr, 2021a; Friston, Da Costa, and Parr, 2021b) have sought to address this by seeking additional conditions or conjectures under which the needed relationship holds. However, the fact that these conditions are highly non-trivial suggests that the scope of the FEP may be much more limited than previously thought.

Furthermore, (1) itself puts a very strong constraint on a system's dynamics. One way to see this is via the *data processing inequality* (Cover & Thomas, 2006, p. 34), which imposes that if (1) holds then all information that μ_t and ϕ_t share needs to be present in $(s_t,\ a_t)$. This would mean that the internal and external states could share no more information than is contained in the sensor and motor states *at the current time.*

But cases where information is stored in the environment and the agent but not in the blanket are ubiquitous. Imagine a friend gives you a phone number written on a piece of paper, which you memorise and then store in a box. The statistical independence between internal and external variables conditioned on active and sensory ones is broken as soon as the piece of paper is away from your sensory input. Once it's out of sight the phone number cannot be stored simultaneously in your internal state and on the piece of paper. As Parr, Da Costa, Heins, Ramstead, and Friston (2021) discuss, this need not be true in transients even if it holds in stationary state. Nevertheless it puts an unrealistic constraint on the stationary dynamics, which we don't expect to be applicable to living organisms.

A possible resolution of this limitation follows from Figure 1. Although (1) cannot be assumed for a general sensor–motor loop,