**Scientific Realism about Friston blankets without Literalism**

**Authors:**

**Julian Kiverstein1 & Michael Kirchhoff2**

1.Amsterdam University Medical Centre, Department of Psychiatry, The Netherlands.

2. University of Wollongong, School of Liberal Arts, Australia

Emails:

[j.d.kiverstein@amsterdamumc.nl](mailto:j.d.kiverstein@amsterdamumc.nl)

[kirchhoff@uow.edu.au](mailto:kirchhoff@uow.edu.au)

**Conflict of interest (**Kiverstein and Kirchhoff): none

**Acknowledgement of funding support:**

Kiverstein is supported by an Amsterdam Brain and Cognition Project Grant;

The European Research Council (Grant number 679190) and the

Netherlands Scientific Organisation (NWO, Vici, awarded to Erik Rietveld). Kirchhoff is supported by Australian Research Council Discovery Project Minds in Skilled Performance (Grant number DP170102987)

**This paper is a commentary on Bruineberg, J., Dolega, K., Dewhurst, J., and Baltieri, M. (2021). The Emperor’s New Markov Blankets. *Behavioral and Brain Sciences*.**

**Abstract:**

Bruineberg and colleagues' critique of Friston blankets relies on what we call the “literalist fallacy”: the assumption that in order for Friston blankets to represent real boundaries, biological systems must literally possess or instantiate Markov blankets. We argue that it is important to distinguish a realist view of Friston blankets from the literalist view Bruineberg and colleagues critique.

[58 words of 60]

**Commentary:**

In our commentary we set out to offer a defence of scientific realism about Markov blankets. Bruineberg and colleagues are right to highlight the choices that go into constructing a scientific model, such as a causal Bayesian network. However, it doesn’t follow that scientific models thereby do not, or cannot, indirectly represent - even if only in an approximate fashion - target systems. Second, we argue that Bruineberg and colleagues' critique of the FEP relies on a fallacy we call the “literalist fallacy”. This is the fallacy of assuming that in order for Markov blankets to identify real boundaries, biological systems must literally possess or instantiate Markov blankets. Scientific realism should be distinguished from literalism, as Bruineberg and colleagues acknowledge. In sum, we conclude that while the target article asks many important questions about the use of the Markov blanket construct in the context of the FEP, it falls short of making the case against scientific realism.

Bruineberg and colleagues, in their lucid presentation of the free energy principle (FEP), make a helpful distinction between three meanings that attach to the use of the word “model”. The first use of the term refers to the scientist’s explanatory target: the neurobiological systems that are argued, by proponents of the FEP, to literally *be* models of their environments (Friston 2013). Second, are the models the scientist makes of neurobiological systems by applying the mathematics of the FEP. Finally, following a significant amount of idealisation and abstraction on the part of the scientist, the scientist arrives at an explanatory model that purports to represent something of interest about a target system. It is at this stage in the modelling process that we arrive at the Friston blanket: a mathematical construct that purports to describe the autopoietic processes that produce a boundary separating the agent from its environment (Kirchhoff et al. 2018).

Bruineberg and colleagues correctly argue that to identify a Friston blanket certain nodes in a Bayesian causal network have to be labelled as internal, external, active and sensory states. They go on to argue (in section 5) that the location of the Markov blanket within a model is largely dependent on the “arbitrary” modelling choices of the scientist. The Markov blanket construct should therefore be understood as a property of the model the scientist is constructing. To think otherwise is to “reify” the Markov blanket, mistaking a construct that is the outcome of modelling statistical relationships of conditional independence, for the cause of this conditional independence.

We fully agree with Bruineberg et al., that the scientist has to make choices about where to locate the Markov blanket within a model. It does not follow, as Bruineberg and colleagues claim, that such choices are arbitrary. The scientist’s decisions about how to interpret a model are a part of the process of model-building. The Markov blanket is a formal or mathematical construct. To model anything this mathematical construct has to be given an interpretation by the scientist. This interpretation does however purport to represent (i.e. describe and explain), the unobservable causes of the behaviour of real world target systems. In the case of Markov blankets, the unobservable causes are the autopoietic processes that produce and maintain a boundary distinguishing the individual agent from its environment. The choices the scientist makes about how to interpret the Markov blanket formalism are therefore not arbitrary. They are guided by the explanatory interests of the scientist, which in this case concern the process of autopoiesis.

Bruineberg and colleagues devote a good deal of attention to uncovering hidden assumptions that are required to apply Friston blankets to biological systems. Scientific models are however very often idealised models that allow for highly complex and intractable problems to be solved - e.g., placing a free energy bound on entropy. One might think that idealisation rules out models from providing accurate representations of their target systems. Idealisation introduces distortion into a model, rendering the resulting model inaccurate. Such an objection rests on a short term view of what scientific modelling can contribute. As Weisberg (2007) notes, scientific idealisation is best understood in the context of a longer-term scientific programme to provide an accurate representation of a target system. A scientific model can represent a target system partially, approximately or probably. The descriptions of the system the scientist provides need not be literally true of the system to approximately describe the behaviour of a target system.

Consider again, with these helpful reminders from the philosophy of science in place, the philosophical mistake Bruineberg and colleagues claim to have uncovered in the very idea of Friston blankets. We have argued that Friston blankets are interpretations given of the Markov blanket formalism in the context of the FEP that purport to describe autopoietic processes. Bruineberg and colleagues claim that to take Friston blankets to represent the processes that cause the conditional independencies differentiating the agent from its environment, one must take the biological agent to literally instantiate or possess a Markov blanket. However, such a claim relies on a fallacious assumption: that scientific realism implies literalism. The realist claims that the Markov blanket formalism can, as an idealised interpretation of a model, nevertheless purport to represent an unobservable causal property. The literalist claims that for a model to represent an unobservable cause, the model must literally be true of a system. Bruineberg and colleagues’ argument against Friston blankets trades on a confusion of realism and literalism we call the “literalist fallacy”.

The target article offers other arguments against Friston blankets. They suggest for instance that for a Friston blanket to mark the boundary of a biological system one must already know where to place the boundary. However, this neglects the application of Friston blankets to extended cognitive systems whose boundaries are negotiable (Clark 2017; Kirchhoff & Kiverstein 2021). The question of where to place the boundary of the system is what is at stake in applying the Markov blanket formalism to such systems (Kirchhoff & Kiverstein 2019). Bruineberg and colleagues have not shown that Friston blankets cannot help to settle such a question.

[1005 words]

**References**

Clark, A. (2017). How to knit your own Markov blanket: Resisting the second law with metamorphic minds. In T. Metzinger & W. Wiese (Ed’s) Philosophy and Predictive Processing: 3. (Frankfurt am Main: MIND Group) doi: 10.15502/9783958573031

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

Kirchhoff, M.D. & Kiverstein, J. (2021). How to determine the boundaries of the mind: a markov blanket proposal. Synthese, https://doi.org/10.1007/s11229-019-02370-y.

Kirchhoff, M. D., Parr, T., Palacios, E., Friston, K. J., and Kiverstein, J. (2018). The markov blankets of life: autonomy, active inference and the free energy principle. Journal of The Royal Society Interface, 15(138): 20170792.

Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy*, 104(102), 639-659.