Markov blankets: Realism and our ontological commitments

Danielle J. Williams

Department of Philosophy, University of California at Davis, Davis, CA 95616, USA
heywilliams@ucdavis.edu
daniellejwilliams.com

doi:10.1017/S0140525X22000255, e217

Abstract

The authors argue that their target is orthogonal to the realism and instrumentalist debate. I argue that it is born directly from it. While the distinction is helpful in illuminating how some ontological commitments demand a theory of implementation, it's less clear whether different views cleanly map onto the epistemic and metaphysical uses defined in the paper.

Bruineberg and colleagues argue there is a conflation between two uses of Markov blankets. Some use Markov blankets in an epistemic way while others use them to make ontological claims about the physical world. To solve this conflation, they propose that we should classify the former as Pearl blankets and the latter as Friston blankets. While this strategy provides a helpful labeling scheme for different uses, a need for a distinction of this kind is indicative of a more substantial problem. Thus, solving this conflation targets a symptom of a broader problem rather than targeting what is at issue in the first place. The authors note that their discussion is orthogonal to the realism and instrumentalism debate in cognitive science, but I argue that their distinction is better understood as a case study born directly from this debate. Computational models play different roles in our scientific theories. We can understand them as purely formal, or we can take them as literally representing physical systems. But, regardless of our position, we need to say something about how our formal, non-physical models relate to the concrete, physical world.

Pearl blankets are Markov blankets used in the formal sense while Friston blankets are taken to be or to genuinely represent concrete boundaries. This distinction rests on how scientists use Markov blankets in their theorizing. But distinguishing between uses leads to a question of how we should frame the difference between Pearl and Friston blankets as scientific posits, not just how they are used within a theory. We could understand the distinction most straightforwardly as delineating between the formal and the physical. One way to cash this out is by thinking about Markov blankets at either the algorithmic level or the implementation level within the Marrian framework. Pearl blankets are purely formal models at the algorithmic level deployed irrespective of the nuts and bolts of the physical system while Friston blankets are implementations of Markov blankets themselves. Because realism proposes that our best scientific theories provide us with knowledge of the objective world – which ontologically commits us to the entities they posit – Markov blankets understood at the implementation level are a bona fide example of a realist position while Markov blankets understood at the algorithmic level and deployed in the Pearl sense demonstrate an instrumentalist position. Because of this, the distinction is not orthogonal to the realism and instrumentalist debate: it's a case study within it.

The authors argue that Friston users have an additional explanatory task because we can't simply read our ontology off of the mathematics. What is needed is an explanation of how a formal construct can be understood in a such metaphysically laden way. This is exactly correct: To complete the theory an account of implementation is required. What is needed for proper reification is an account that maps the formal mathematical model to the boundaries of the physical world. While it is still an open question how we should formulate the implementation relation, there are some views that could be adopted. One approach is to argue that there must be some resemblance between the model and the target system such that some specified features are necessarily consistent between the two (Curtis-Trudel, 2021). Resemblance may help to alleviate some conceptual issues regarding irregular boundaries. Another viable option comes from Bogacz (2015). Bogacz proposes a theory of implementation that maps different elements of the model onto different neural populations within the cortex where the mapping between the variables in the model and the elements of the neural circuitry may not be "clean" but rather "messy" (Bogacz, p. 209). Different views will map the formal computation onto the physical world in different ways, but what is important is that the relation between the formal model and the physical world is accounted for.

One worry, though, is that the distinction between Pearl blankets and Friston blankets is overly restrictive. There are additional ways to understand how Markov blankets are used over and above the Pearl and Friston senses. For example, one might be a realist without being committed to physical implementation: It is possible to have ontological commitments to mathematical entities at Marr's algorithmic level without ontologically committing oneself to implementation level features. Scientific realism proposes that we are ontologically committed to the existence of the posits that do explanatory work in our best scientific theories. Depending on your view of explanation, non-causal, formal properties can play a robust explanatory role that meets the criterion for scientific realism (Williams & Drayson, forthcoming). This goes beyond the epistemic use and stops just short of the metaphysical use blurring the distinction between Pearl and Friston blankets by neglecting to carve out space for a mathematical ontology. If one can hold ontological commitments about formal entities, do they also have an additional explanatory debt? Do they now count as Friston blankets? Because you can have ontological commitments at both the formal and physical levels, the distinction between Pearl and Friston uses blurs and additional explanatory requirements become unclear.

Different uses of Markov blankets provide a case study within the instrumentalism and realism debate in cognitive science. Some accept the formal model as an epistemic tool while others use the formal model to make ontological claims. As with all formal models, for proper reification, some account of implementation is needed. But, once the distinction is considered within the context of the realism and instrumentalism debate in which it belongs, it become unclear that the distinction is able to do the work that it sets out to do in the first place because it fails to leave room for additional ways in which one can take on a realist stance about formal models.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

References

Bogacz, R. (2015). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematics and Psychology*, *76*, 198–122.

Curtis-Trudel, A. (2021). Implementation as resemblance. *Philosophy of Science*, *88*, 1021–1032.

Williams, D., & Drayson, Z. (forthcoming). The nature of the predictive mind: Realism and instrumentalism in Bayesian cognitive science. In T. Cheng, R. Sato & J. Hohwy (eds.) *Expected experiences: The predictive mind in an uncertain world*. Routledge.

Causal surgery under a Markov blanket

Daniel Yon[a] and Philip Robert Corlett[b]

[a]Department of Psychological Sciences, Birkbeck, University of London, London WC1E 7HX, UK and [b]Department of Psychiatry, Connecticut Mental Health Center, Yale University, New Haven, CT 06519, USA
d.yon@bbk.ac.uk
philip.corlett@yale.edu
www.danielyon.com
https://medicine.yale.edu/lab/corlett/

doi:10.1017/S0140525X22000218, e218

Abstract

Bruineberg et al. provide compelling clarity on the roles Markov blankets could (and perhaps should) play in the study of life and mind. However, here we draw attention to a further role blankets might play: as a hypothesis about cognition itself. People and other animals may use blanket-like representations to model the boundary between themselves and their worlds.

In their impressive target article Bruineberg et al. describe two radically different ways we can use Markov blankets. *Pearl blankets* are tools that allow scientists to identify (in)dependence between variables when modelling complex systems. In contrast, *Friston blankets* are tools philosophers may use to parse the world into internal and external states, distinguishing agents from the rest of their worlds.

We wholeheartedly agree that this distinction is important, but feel this dichotomy neglects a third possibility: blankets as a hypothesis about cognition itself. In this way of thinking, cognising creatures may use processes that approximate Bayesian modelling to track which states of the world depend on or are independent of their actions. In so doing, these creatures construct a *cognitive blanket* that captures their beliefs about what they can and cannot control.

This *cognitive blanket* hypothesis makes distinctive predictions about how agents estimate agency and control over their bodies and the world. Many have suggested that humans and other animals determine what they can control by tracking correlations between actions and outcomes (Dickinson & Balleine, 1994; Yon, Bunce, & Press, 2020). However, building a cognitive blanket – mapping causal dependencies between actions and states – allows an agent to entertain counterfactual scenarios and to intervene on the world to test connections implied by their model. This kind of hypothesis testing – evocatively dubbed "causal surgery" (Pearl, Glymour, & Jewell, 2016) – allows agents to refine beliefs about their own causal power by acting on the world in informative ways.

Psychologists can test for *cognitive blankets* by investigating whether agents are sensitive to counterfactual information and engage in "causal surgery" to test what they can and cannot control. In humans, there is some evidence of sensitivity to counterfactual information – we feel a greater sense of control when we believe we could have acted differently and this could have altered outcomes (Kulakova, Khalighinejad, & Haggard, 2017). There is also tentative evidence that human agents perform exploratory actions when judging control over events in the external world (Wen et al., 2020). This kind of exploration could be a hallmark of "causal surgery" that tests hypotheses about our influence. However, it is also possible that apparently exploratory behaviour emerges from noise in decision and action systems (Findling, Skvortsova, Dromnelle, Palminteri, & Wyart, 2019). Targeted tests are thus needed to establish whether humans engage in genuine causal surgery when estimating control – possibly by determining whether explorations about control depend on the agent's uncertainty about action–outcome relationships.

The same tests could also be applied by comparative cognitive scientists. It has long been debated how far nonhuman animals represent their behaviour as "causes" of environmental changes (Penn & Povinelli, 2007). In our way of thinking, empirical evidence of causal surgery in different species would suggest the animal is constructing a *cognitive blanket* – testing hypotheses about how action and outcome connect. As with humans, it would be important to distinguish uncertainty-driven hypothesis testing in animals from blind exploration. Such efforts could exploit apparent signatures of "confidence" detectable in animals (Kepecs, Uchida, Zariwala, & Mainen, 2008), or could investigate how animals respond to different varieties of environmental uncertainty (Yon & Frith, 2021). For instance, if a creature's exploratory behaviour responds to volatility in action–outcome relationships, this may be indicative of causal surgery: The active probing of the agent's blanket-like model to test what they can and cannot influence.

Furthermore, *cognitive blankets* could illuminate the disturbances of action awareness that occur in psychiatric illness. Patients with psychosis often develop delusions about action and control: They claim to control things they objectively cannot (grandiosity) and deny controlling some actions they have genuinely authored (passivity; Frith, Blakemore, and Wolpert, 2000). These strange beliefs might arise from a disordered blanket that draws the boundary between world and agent in an unusual way (much like that depicted in Bruineberg et al.'s Fig. 7c). If intervention and exploration are essential ingredients in building up an accurate *cognitive blanket*, it may be fruitful for clinical scientists to investigate processes of causal surgery in psychosis. If these patients are less likely to intervene on the world to test what they can control, unusual beliefs about the self and the world may persist unchecked. Indeed, one could speculate that a vicious cycle obtains in psychosis, where negative symptoms dampening the drive to act (e.g., apathy, catatonia) rob patients of action–outcome experiences that could challenge positive symptoms (i.e., delusions about action; see Bortolotti & Broome, 2012; Corlett, Honey, & Fletcher, 2016). We note with interest the role that dopamine signalling appears to play in learning, confidence, causal inference, and their derangement in psychosis-like states (Redgrave & Gurney, 2006; Schmack, Bosc, Ott, Sturgill, & Kepecs, 2021; Sharpe et al., 2017).

Our third way of thinking about blankets – as representations in the heads of agents – departs from both *Friston* and *Pearl*

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