"The Emperor's New Markov Blankets" (Bruineberg, Dolega, Dewhurst, Baltieri)
Commentary

WORD COUNTS

Abstract:	52
Main text:	997
References:	157
Entire text:	1291

TITLE: **Bayesian realism and structural representation**

Alex Kiefer
Cognition & Philosophy Lab
Monash University
20 Chancellors Walk, Room E674
Monash University VIC 3800
Australia
Ph. +61 3 99053208
Alex.Kiefer@monash.edu
alexbkiefer.net

Jakob Hohwy
Monash Center for Consciousness & Contemplative Studies
Cognition & Philosophy Lab
Monash University
20 Chancellors Walk, Room E674
Monash University VIC 3800
Australia
Ph. +61 3 99053208
Jakob.Hohwy@monash.edu
https://research.monash.edu/en/persons/jakob-hohwy

ABSTRACT: We challenge the authors' view that Markov blankets are illicitly reified when used to describe organismic boundaries. We do this both on general methodological grounds, where we appeal to a form of structural realism derived from Bayesian cognitive science to dissolve the problem, and by rebutting specific arguments in the target article.

MAIN TEXT

In this commentary we argue that, from the point of view of Bayesian cognitive science, concerns about the reification of Markov blankets are misplaced. We assume that scientific theories represent reality in the same way in which we have argued (Kiefer and Hohwy 2017; Kiefer and Hohwy 2019) that organisms' internal generative models represent their external worlds: by way of (exploitable) structural similarity. Mathematically articulated theories or models set up relations among their variables which, when successful, mimic the relations that obtain among elements of the target system (Cummins 1991). In asserting such theories, we impute these very structural properties to the world.

This view is fundamentally a structural-realist one (Ladyman, 1996; Ladyman and Ross, 2007), and is in a sense nothing new, but whether novel or not it is adequate to the task of addressing the present criticism. To start, this perspective on the semantics of scientific theories completely defuses the concern that projection of the content of a mathematically expressed theory (such as the FEP) onto the physical world involves a category-mistake or fallacy. The "substance" of the representational medium (whether collections of abstract mathematical symbols or bits of pasta) makes no difference apart from its expressive adequacy, since it is the form and not the substance of the target domain that the theory attempts to capture in virtue of *its* form.

A traditional argument against instrumentalism (Putnam 1975) has it that theories are predictive only to the extent that they are true. Structuralist representationalism allows us to nuance this argument: we may expect observational adequacy *to the degree* that the structure of the theory matches that of the generator of observations. A Bayesian might point out that the confidence we (ought to) place in a theory also scales with its predictive accuracy (as well as its prior plausibility). By adopting a Bayesian attitude toward scientific theories, we can then dispense with a categorical distinction between realism and instrumentalism, while taking on board the epistemic humility that makes the latter appealing.

The preceding remarks do not on their own, of course, justify a realist attitude toward Markov blankets in living systems or elsewhere—this would depend on the empirical adequacy of such Markovian descriptions. They do however suggest that there is nothing *methodologically* flawed in the practice of imputing formally characterized structures (such as structures of conditional independence in Bayesian networks) to the real world.

In the remainder we rebut three arguments given in the target article that purport to establish the contrary.

The first argument poses an analogy between Markov blankets and contour lines on cartographic maps. I expect to observe rivers and mountains when I navigate by a map, but *pace* the authors, I also expect to experience elevation and other aspects of the terrain represented by contour lines. To mistake contour lines as such for features of the terrain would indeed be a radical mistake, but so would expecting the river to be ink-blue. We do not make

these mistakes in practice any more than proponents of the FEP suppose on the basis of their diagrams that the sensory epithelia of organisms consist in labelled circles with black outlines.

The moral is that contour lines contribute to the same fundamentally spatial (indeed structural-similarity-based) representation as other elements of the map, though in a slightly more abstract and conventional way. There may be more reasonable concerns about reification, for example that graphical models cut the world artificially into discrete, repeatable event-types, but this worry would impugn the use of such abstractions to represent causal structure (such as smoking's causing cancer) quite generally, and so is not properly directed against the FEP literature.

In fact, the authors *do* argue for a blanket instrumentalism with respect to Bayesian networks, which brings us to the second argument: that the choice of model for a given system depends in part on extrinsic circumstances like data availability or the interests of the scientist. But the intrusion of this pragmatic element means merely that there exist *many* sets of conditional dependencies, some more interesting than others, which does not impugn a realist attitude toward any one of those sets.

The authors further suggest that the Markov blanket formalism does little work in delineating organismic boundaries if we must already have selected a model in order to consider its blanket. It is unclear to us how this epistemic point could count against realism, but in any case the plurality of Markov blankets present in any system of interest has long been admitted by FEP theorists (cf. Hohwy 2016; Friston et al 2021), and in practice the FEP is interesting not as a tool for distinguishing organisms from their environments, which we can do well enough without it, but for its formal account of how systems act to *maintain* the integrity of their boundaries, however initially identified.

The final argument we consider suggests that the literature around the FEP stratifies into two distinct projects, the respectable empirical one of using Markov blankets to characterize aspects of organisms' cognitive models of their environments ("inference with a model"), and the metaphysically ambitious one of using Markov blankets to characterize organisms themselves and their boundaries with respect to their environments ("inference within a model"). Here, it is sufficient to point out that these projects are not, actually, fundamentally distinct in kind. The explanatory targets in both cases are real features of organisms (their cognitive models, in the first case, and their sensorimotor boundaries, in the second). What may be controversial in the case of the FEP is the idea that the entire organism (as opposed to some construct in its brain) may be regarded as a "model", but this is not the ground on which the authors stake their claim.

In conclusion, we have seen no reason, either on the basis of its general mathematical character or on the basis of its particular modes of application, to suspect that the Markov blanket formalism has been used by FEP theorists to commit fallacies of reification.

CONFLICT OF INTEREST STATEMENT

Conflicts of interest: none.

FUNDING STATEMENT

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

References

Cummins, R. (1991). *Meaning and Mental Representation*. Cambridge: MIT Press.

Friston, K. J., Fagerholm, E. D., Zarghami, T. S., Parr, T., Hipólito, I., Magrou, L., & Razi, A. (2021). Parcels and particles: Markov blankets in the brain. *Network Neuroscience*, *5*(1), 211–251. https://doi.org/10.1162/netn_a_00175

Hohwy, J. (2016). The self-evidencing brain. *Noûs*, *50*(2), 259–285. https://doi.org/10.1111/nous.12062

Kiefer, A.B. and Hohwy, J. (2017). Content and misrepresentation in hierarchical generative models. *Synthese* 195. Special issue on predictive brains (M. Kirchhoff, ed.).

Kiefer. A.B. and Hohwy, J. (2019). Representation in the Prediction Error Minimization Framework. In S. Robins, J. Symons, & P. Calvo (eds.), *The Routledge Companion to the Philosophy of Psychology: 2nd Edition*. London: Routledge.

Ladyman, James, 1998, What is Structural Realism?, Studies in History and Philosophy of *Science*, 29(3): 409–424. doi:10.1016/S0039-3681(98)80129-5

Ladyman, James & Don Ross, 2007, *Every Thing Must Go: Metaphysics Naturalised*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199276196.001.0001

Putnam, H. (1975). *Mathematics, matter and method*. Cambridge: Cambridge University Press.