the dynamics. As a consequence, neither this result nor blankets themselves would follow from first principles, but rather from a fallible heuristic analysis of the system of interest.

On the other hand, if we eschew precision and accept approximate optimization methods for finding Pearl blankets such as those widespread in machine learning and causal search (e.g., Pellet & Elisseeff, 2008), we can use them as tools of discovery to identify the boundaries of entities (e.g., nodes in neural networks for the purpose of systems neuroscience). Furthermore, showing that a system delineated in this way conforms to the FEP might provide much more insight into the nature of the process, as it would require less knowledge at the outset. However, approximate methods do not allow for the use of the concept of Friston blanket and effectively preclude the viability of the metaphysical programme of the FEP as a naturalist ontology for life sciences.

Perhaps it is too quick to throw the blankets entirely at this point. Nonetheless, we believe that the use of the Markov blanket construct should enable us to solve pressing issues in computational modelling in the sciences of brain and behaviour. While Markovian monism metaphysics is not such a pressing issue, studying the causal and functional dynamics of cognitive systems is. In this context, we need various fallible heuristics for delineating Pearl blankets; that is, many stupid (Smaldino, 2017), approximate, and tractable models, and we need more of them to be able to make use of the error diversity inherent in any heuristic enterprise (Wimsatt, 2007). While stronger analytical methods for finding Markov (and Friston) blankets are not necessarily dead ends, the FEP theorists' focus on those difficult methods makes them overlook a lot of lower hanging fruits.

Acknowledgements. We want to thank Mel Andrews, Conor Heins, Dalton Sakthivadivel, and the Active Inference Lab for helpful clarifications.

Financial support. This work was supported by the Ministry of Education and Science (Poland) research grant (W. R., DI2018 010448), as part of the "Diamentowy Grant" programme; National Science Centre (Poland) research grant (M. M., P. L., grant number 2014/14/E/HS1/00803) and Leverhulme Doctoral Scholarship (T. K.).

Conflict of interest. None.

References

Aliferis, C. F., Tsamardinos, I., & Statnikov, A. (2003). HITON: a novel Markov blanket algorithm for optimal variable selection. *AMIA 2003 Annual Symposium Proceedings. AMIA Symposium*, 2003, 21–25.

Bai, X., Glymour, C., Padman, R., Ramsey, J., Spirtes, P. L., & Wimberly, F. C. (2004). *PCX: Markov blanket classification for large data sets with few cases*. Center for Automated Learning and Discovery. Retrieved from http://reports-archive.adm.cs.cmu.edu/anon/cald/CMU-CALD-04-102.pdf

Da Costa, L., Friston, K., Heins, C., & Pavliotis, G. A. (2021). Bayesian mechanics for stationary processes. Proceedings of the Royal Society A: Mathematical, *Physical and Engineering Sciences*, *477*(2256), 20210518. https://doi.org/10.1098/rspa.2021.0518

Friston, K., Heins, C., Ueltzhöffer, K., Da Costa, L., & Parr, T. (2021a). Stochastic chaos and Markov blankets. *Entropy*, *23*(9), 1220. https://doi.org/10.3390/e23091220

Friston, K. J. (2019). A free energy principle for a particular physics. *arXiv:1906.10184 [q-bio]*. http://arxiv.org/abs/1906.10184

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

Friston, K. J., Wiese, W., & Hobson, J. A. (2020). Sentience and the origins of consciousness: From Cartesian duality to Markovian monism. *Entropy*, *22*(5), 516. https://doi.org/10.3390/e22050516

Pellet, J.-P., & Elisseeff, A. (2008). Using Markov blankets for causal structure learning. *Journal of Machine Learning Research*, *9*(43), 1295–1342.

Peña, J. M., Nilsson, R., Björkegren, J., & Tegnér, J. (2007). Towards scalable and data efficient learning of Markov boundaries. *Eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, 45(2), 211–232. Retrieved from https://doi.org/10.1016/j.ijar.2006.06.008

Smaldino, P. E. (2017). Models are stupid, and we need more of them. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational social psychology* (pp. 311–331). Routledge. https://doi.org/10.4324/9781315173726-14

Tsamardinos, I., Aliferis, C. F., & Statnikov, A. (2003). Time and sample efficient discovery of Markov blankets and direct causal relations. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 673–678. https://doi.org/10.1145/956750.956838

Wiese, W., & Friston, K. J. (2021). Examining the continuity between life and mind: Is there a continuity between autopoietic intentionality and representationality? *Philosophies*, *6*(1), 18. https://doi.org/10.3390/philosophies6010018

Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Harvard University Press.

Markov blankets as boundary conditions: Sweeping dirt under the rug still cleans the house

Javier Sánchez-Cañizares

Mind-Brain Group at the Institute for Culture and Society (ICS), University of Navarra, 31009 Pamplona, Spain
js.canizares@unav.es
https://www.issr.org.uk/fellows/user/496/

doi:10.1017/S0140525X22000097, e207

Abstract

Bruineberg et al. underestimate the ontological weight of Markov blankets as actual boundaries of systems and lean toward an instrumentalist understanding thereof. Yet Markov blankets need not be deemed mere tools. Determining their reality depends on the fundamental problem of distinguishing between system and environment in physics, which, in turn, demands a metaphysical bedrock backed by a realist stance on science.

Do Markov blankets have any ontological weight? Most of the literature on the free energy principle (FEP) equates Markov blankets with actual system boundaries without further ado. However, as the authors rightly point out, there is a crucial difference between using blankets as an epistemic tool in Bayesian networks to identify independencies among random variables (Pearl blankets) and using them as actual boundaries between a system and its environment (Friston blankets).

I agree with the authors on the insufficiency of Friston blankets to demarcate the physical boundary between an agent and its environment, and the need for additional philosophical assumptions "to do such heavy metaphysical lifting." FEP theorists seem aware of this since they recently gave their framework a freshly minted metaphysical interpretation, dubbed Markovian monism (MM) (Friston, Wiese, & Hobson, 2020). According to MM, the very fact that one may demarcate a system from its environment through a Markov blanket induces a dual aspect information geometry in the system's internal states that enables it to represent its surroundings. MM thus reveals the troublesome transition between Markov blankets as epistemic model-bound

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

tools and their alleged ontological consequences, which are beset by circular reasoning (Sánchez-Cañizares, 2021).

However, within FEP formalism, the ontology proper to Markov blankets need not take the brunt of the blow because it inherits the more general problem of distinguishing between a system and its environment in physics. In the absence of a theory of everything, conventional physics must accept initial conditions to start computing the world's behavior. "You need a starting point!" (Wilczek, 2015). Remarkably, the existence of systems seems to require the universe to have exceptional initial conditions and dynamics (Tegmark, 2015) since the theory that is currently most fundamental, that is, quantum mechanics, deems the distinction between system and environment as relative (Lombardi, Fortín, & Castagnino, 2012). Physics teaches us that one must make an *ansatz* to progress in the scientific description of nature. Whether a chosen *ansatz* holds at a specific process only becomes evident *a posteriori*. Markov blankets are undoubtedly the most basic *ansatz* for FEP formalism to work.

Things being so, the authors' critiques of Markov blankets as Friston blankets become less weighty in two interrelated respects:

(1) Markov blankets define the dominion – via the partition of variables – for which it makes sense to minimize free energy. Their precise definition must change whenever the model becomes unsuitable for describing unexpected dynamics, that is, dynamics that cannot be fully grasped within the assumptions of a particular model. But this does not differ from the usual procedure of changing boundary conditions for the distinct models that are compatible with a principle theory. The authors themselves recognize this when quoting Andy Clark, "boundaries are malleable (over time) and multiple." FEP is a principle and needs reinterpretation for each model. One such reinterpretation must state what the system and its environment are, even though such a distinction does not strictly stem from FEP formalism. As a consequence, one should not consider FEP as wholly explanatory of living beings' natural history (Longo & Montévil, 2014), if only because Friston blankets also change throughout the system's history.

(2) Even though Friston blankets are metaphysically wanting, they should not be judged as a mere instrumental tool. The authors, however, ultimately lean toward "a strongly instrumentalist understanding of Bayesian networks, and hence of Markov blankets, which would not justify the kinds of strong philosophical conclusions drawn by some from the idea of a Friston blanket." Instrumentalism undoubtedly looms large in philosophical interpretations of scientific research, but, in the end, this seems to be a self-defeating strategy for several reasons:

    (2.1) The instrumentalist that deems Markov blankets as Pearl blankets forgoes endowing the former with any ontological weight. Yet such a mindset could easily lead to denying the very existence of systems – as sheer constructs of human perception. Nevertheless, if one does not wish to embrace such a radical position and, on the contrary, accepts the existence of systems in the universe, something quite similar to Friston blankets must also exist in each system in order to sieve through the many environmental influences that foster or threaten the system's identity.

    (2.2) It seems paradoxical to emphasize the insufficient justification for transforming Pearl blankets into Friston blankets whereas, ultimately, glossing over what additional philosophical assumptions might look like for such a move. A scientific realist, for instance, could call on formal causation as a valid metaphysical framework that allows for the use of ad hoc boundary conditions to individuate systems that enjoy specific dynamics in nature (Owen, 2020, 2021; Sánchez-Cañizares, 2022a, 2022b). In other words, Markov blankets reflect the emergence of boundary conditions for living systems. If boundary conditions exist for some relevant time scale, Markov blankets are Friston blankets. In addition, such a philosophical commitment can adequately frame the emergence of complex dynamical systems, which one cannot just deduce from their underlying dynamics (Juarrero, 2002; Sánchez-Cañizares, 2016).

    (2.3) The methodological issue at play refers to whether other kinds of knowledge – for example, knowing living systems as wholes – that influence scientific research should be accepted within the overall explanatory picture. Certain pre-scientific knowledge is necessary for guiding scientific methodology. If one assumes said knowledge, there is no fundamental reason to deny that some Markov blankets are also Friston blankets, even if for a limited period, or that Friston blankets are not fixed and may transition in a variety of ways toward different instantiations. In doing so, the inevitable, closed circularity of scientific instrumentalism turns into the open circularity of scientific realism, which admits a hierarchical variety of assumptions and hypotheses about the world, as well as the possibility of cognitive progress based on constant confrontation with observation.

Financial support. This work has been funded with the help of Fundación Ciudadanía y Valores (FUNCIVA) and Proeduca Summa S.L.

Conflict of interest. None.

References

Friston, K. J., Wiese, W., & Hobson, J. A. (2020). Sentience and the origins of consciousness: From Cartesian duality to Markovian monism. *Entropy*, *22*(5), 516. https://doi.org/10.3390/E22050516

Juarrero, A. (2002). Complex dynamical systems and the problem of identity. *Emergence*, *4*(1), 94–104.

Lombardi, O., Fortín, S., & Castagnino, M. (2012). The problem of identifying the system and the environment in the phenomenon of decoherence. In H. W. de Regt, S. Hartmann & S. Okasha (Eds.), *The European philosophy of science association proceedings: Amsterdam 2009* (pp. 161–174). Springer. https://doi.org/10.1007/978-94-007-2404-4_15

Longo, G., & Montévil, M. (2014). *Perspectives on organisms: Biological time, symmetries and singularities*. Springer.

Owen, M. (2020). Aristotelian causation and neural correlates of consciousness. *Topoi*, *39*(5), 1113–1124. https://doi.org/10.1007/s11245-018-9606-9

Owen, M. (2021). Circumnavigating the causal pairing problem with hylomorphism and the integrated information theory of consciousness. *Synthese*, *198*, 2829–2851. https://doi.org/10.1007/s11229-019-02403-6

Sánchez-Cañizares, J. (2016). Entropy, quantum mechanics, and information in complex systems: A plea for ontological pluralism. *European Journal of Science and Theology*, *12*(1), 17–37.

Sánchez-Cañizares, J. (2021). The free energy principle: Good science and questionable philosophy in a grand unifying theory. *Entropy*, *23*(2), 238. https://doi.org/10.3390/e23020238

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

Sánchez-Cañizares, J. (2022a). Formal causation in integrated information theory: An answer to the intrinsicality problem. *Foundations of Science*, *27*, 77–94. https://doi.org/10.1007/s10699-020-09775-w
Sánchez-Cañizares, J. (2022b). Integrated information theory as testing ground for causation: Why nested hylomorphism overcomes physicalism and panpsychism. *Journal of Consciousness Studies*, *29*(1–2), 56–78. https://doi.org/10.53765/20512201.29.1.056
Tegmark, M. (2015). Consciousness as a state of matter. *Chaos, Solitons and Fractals*, *76*, 238–270. https://doi.org/10.1016/j.chaos.2015.03.014
Wilczek, F. (2015). *A beautiful question. Finding nature's deep design*. Penguin Press.

A continuity of Markov blanket interpretations under the free-energy principle

Anil Seth, Tomasz Korbak and Alexander Tschantz

Department of Informatics, School of Engineering and Informatics, University of Sussex, Brighton BN1 9QJ, UK
a.k.seth@sussex.ac.uk
tomasz.korbak@gmail.com
tschantz.alec@gmail.com

doi:10.1017/S0140525X2200036X, e208

Abstract

Bruineberg and colleagues helpfully distinguish between instrumental and ontological interpretations of Markov blankets, exposing the dangers of using the former to make claims about the latter. However, proposing a sharp distinction neglects the value of recognising a continuum spanning from instrumental to ontological. This value extends to the related distinction between "being" and "having" a model.

"We should not confuse the foundations of the real world with the intellectual props that serve to evoke that world on the stage of our thoughts." This quote from Ernst Mach (Mach [2012], p. 531, translated in Sigmund [2017], p. 19), surfacing from the origins of the philosophy of science, connects directly to the target article, in which Bruineberg and colleagues discuss how Markov blankets (MBs) should be understood within the wider literature of the free energy principle (FEP, Friston, 2010), as well as how "models" and "modelling" should be interpreted within the cognitive and brain sciences more generally.

MBs are statistical descriptions that partition systems into internal, external, and blanket variables – where the internal variables are conditionally independent of the external variables, given the blanket variables. Bruineberg et al. provide a valuable service by distinguishing two interpretations of MBs: "Pearl blankets" (PBs) and "Friston blankets" (FBs). PBs embody an instrumental approach, in which MBs are used as tools to aid the analysis of complex systems, for example by identifying sets of variables suitable for further investigation. In contrast, FBs adopt an ontological stance in which they are assumed to either *be* (a literalist reading) – or *usefully approximate* (a realist reading) actually existing boundaries in the world, such as the boundary between a cell and its milieu, or between an organism or agent and its environment. Bruineberg et al. reveal the dangers of conflating these two interpretations, in particular when an instrumental (PB) application is implicitly or explicitly taken to justify ontological (FB) conclusions. Their arguments should be borne in mind by those inclined to help themselves to the FEP to explain their favourite grand mystery, or to take it as gospel.

Having said this, making a sharp distinction is often a useful prelude to recognising a spectrum of positions, each of which may be useful when assessed on its own merits. We suggest this is the case here. For example, one may begin with an instrumentalist approach and progressively refine and extend the corresponding model so as to make increasingly specific claims about the causal mechanisms at play in the system under study – in this way, gradually moving towards a more ontological or realist stance. What does "refine and extend" mean? It could mean equipping the model with additional features that represent potentially important and context-specific aspects of the relevant boundaries, such as autopoietic (self-producing) processes for biological boundaries (Kirchhoff, Parr, Palacios, Friston, & Kiverstein, 2018; Maturana & Varela, 1980), and embodied and embedded interactions for cognitive boundaries (Clark & Chalmers, 1998; Kirchhoff & Kiverstein, 2021), as well as a recognition of the limited degree to which statistical identification of an MB might generalise to nonequilibrium systems (Aguilera, Millidge, Tschantz, & Buckley, 2022; Biehl, Pollock, & Kanai, 2021).

Bruineberg et al. mention these possibilities, but downplay their significance by drawing a contrast between "additional philosophical assumptions" and "additional technical assumptions," where the latter implicitly subsumes everything just mentioned. But these modelling strategies and mathematical constraints are more than just additional assumptions, they can often be part-and-parcel of the explanatory model itself. And rather than "additional philosophical assumptions," what seems to be required is a *recognition* of the model's philosophical context and the claims made on its behalf, so as to avoid the sort of conflation helpfully identified by Bruineberg.

Bruineberg et al. worry that, given such additional aspects, whether the MB formalism itself can still be doing any work? The answer is yes, to the extent that it helps specify those parts of a model that are focused on boundaries. By casting the distinction between PBs and FBs as sharp, rather than as extrema on a continuum, Bruineberg et al. underestimate the explanatory work that MBs may uniquely be able to do.

Digging a little deeper, one reason we might be tempted to invoke a bright-line distinction between PBs and FBs is because of the dramatic claims made for literalist readings of FBs, in which MBs are seen as really existing ontological boundaries in physical systems. But – as Mach reminds us – models are always models, whatever their granularity. Once we discount the relevance of an overly literalist reading, the value of a continuity between instrumental and ontological stances becomes easier to appreciate. (Here, it is worth separating Mach's valuable scepticism about literalism from his ultimately doomed project to ground physics solely in phenomenology; it is not likely that Mach would have had much time for FBs, even of a realist flavour.)

The same reasoning can be applied to the distinction between "being a model" and "having a model" – a distinction that Bruineberg et al. mention, but only in passing (see also Seth and Tsakiris, 2018). Under the FEP, and following the spirit of the cybernetic pioneers (Conant & Ashby, 1970), many systems can be interpreted as "being" a model of their environment. In an example briefly discussed by Bruineberg et al., even a simple Watt governor can be described as performing inference – however it is best thought of not as *having* a model that is used to

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