blankets. Even if *Friston blankets* cannot pick out the objective boundaries between agents and their worlds, blanket-like computations may still be the processes by which some creatures (suboptimally) identify where these boundaries lie.

Moreover, the ability of scientists to model an agent using *Pearl blankets* cannot tell us whether the agent uses a *cognitive blanket* to model itself. For example, creatures that do not build a *cognitive blanket* may rely on simple, lean psychological processes like associative learning to navigate their environments (i.e., without building a causal graph). However, these simpler psychological processes are only adaptive because they also gear creatures into the causal structure of their environments (Papineau & Heyes, 2006). Simple creatures may thus be well modelled by *Pearl blankets*, even if they do not have a *cognitive blanket* of their own.

In conclusion, Bruineberg et al. provide an important perspective on how scientists and philosophers should use Markov blankets to describe the boundaries between agents and their worlds. However, it is also important to consider (and test) whether blanket-like representations are at the heart of how agents construct these boundaries in their own minds.

References

Bortolotti, L., & Broome, M. R. (2012). Affective dimensions of the phenomenon of double bookkeeping in delusions. *Emotion Review*, 4(2), 187–191. https://doi.org/10.1177/1754073911430115

Corlett, P. R., Honey, G. D., & Fletcher, P. C. (2016). Prediction error, ketamine and psychosis: An updated model. *Journal of Psychopharmacology*, 30(11), 1145–1155. https://doi.org/10.1177/0269881116650087

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, 22(1), 1–18. https://doi.org/10.3758/BF03199951

Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., & Wyart, V. (2019). Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nature Neuroscience*, 22(12), 2066–2077. https://doi.org/10.1038/s41593-019-0518-9

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 355(1404), 1771–1788. https://doi.org/10.1098/rstb.2000.0734

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, 455(7210), 227–231. https://doi.org/10.1038/nature07200

Kulakova, E., Khalighinejad, N., & Haggard, P. (2017). I could have done otherwise: Availability of counterfactual comparisons informs the sense of agency. *Consciousness and Cognition*, 49, 237–244. https://doi.org/10.1016/j.concog.2017.01.013

Papineau, D., & Heyes, C. (2006). Rational or associative? Imitation in Japanese quail. In M. Nudds & S. Hurley (Eds.), *Rational animals* (pp. 187–195). Oxford University Press.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. Wiley.

Penn, D. C., & Povinelli, D. J. (2007). Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology*, 58, 97–118. https://doi.org/10.1146/annurev.psych.58.110405.085555

Redgrave, P., & Gurney, K. (2006). The short-latency dopamine signal: A role in discovering novel actions? *Nature Reviews. Neuroscience*, 7(12), 967–975. https://doi.org/10.1038/nrn2022

Schmack, K., Bosc, M., Ott, T., Sturgill, J. F., & Kepecs, A. (2021). Striatal dopamine mediates hallucination-like perception in mice. *Science* (New York, N.Y.), 372(6537), eabf4740. https://doi.org/10.1126/science.abf4740

Sharpe, M. J., Chang, C. Y., Liu, M. A., Batchelor, H. M., Mueller, L. E., Jones, J. L., … Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nature Neuroscience*, 20(5), 735–742. https://doi.org/10.1038/nn.4538

Wen, W., Shibata, H., Ohata, R., Yamashita, A., Asama, H., & Imamizu, H. (2020). The active sensing of control difference. *iScience*, 23(5), 101112. https://doi.org/10.1016/j.isci.2020.101112

Yon, D., Bunce, C., & Press, C. (2020). Illusions of control without delusions of grandeur. *Cognition*, 205, 104429. https://doi.org/10.1016/j.cognition.2020.104429

Yon, D., & Frith, C. D. (2021). Precision and the Bayesian brain. *Current Biology*, 31(17), R1026–R1032. https://doi.org/10.1016/j.cub.2021.07.044

Authors' Response

The Emperor Is Naked: Replies to commentaries on the target article

Jelle Bruineberg[a] [ID], Krzysztof Dołęga[b] [ID],

Joe Dewhurst[c] [ID] and Manuel Baltieri[d,e] [ID]

[a]Department of Philosophy, Macquarie University, Sydney, NSW 2109, Australia; [b]Institut für Philosophie II, Fakultät für Philosophie und Erziehungswissenschaft, Ruhr-Universität Bochum, 44801 Bochum, Germany; [c]Fakultät für Philosophie, Wissenschaftstheorieund Religionswissenschaft, Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität München, 80539 Munich, Germany; [d]Araya, Inc., Tokyo, Japan and [e]School of Engineering and Informatics, University of Sussex, Brighton BN1 9RH, UK
jelle.bruineberg@mq.edu.au
krzysztof.dolega@rub.de
joseph.e.dewhurst@gmail.com
manuel_baltieri@araya.org

Abstract

The 35 commentaries cover a wide range of topics and take many different stances on the issues explored by the target article. We have organised our response to the commentaries around three central questions: Are Friston blankets just Pearl blankets? What ontological and metaphysical commitments are implied by the use of Friston blankets? What kind of explanatory work are Friston blankets capable of? We conclude our reply with a short critical reflection on the indiscriminate use of both Markov blankets and the free energy principle.

R1. Introduction

"What's this?" thought the Emperor. "I can't see anything. This is terrible! Am I a fool? Am I unfit to be the Emperor? What a thing to happen to me of all people! – Oh! It's very pretty," he said. "It has my highest approval." And he nodded approbation at the empty loom. Nothing could make him say that he couldn't see anything.

In H. C. Andersen's folktale *The Emperor's New Clothes*, two swindlers convince the king to let them make him a set of special clothes that is invisible to anyone who is either too stupid or incompetent. One by one, the Emperor's advisors go and check with the swindlers and, afraid of being thought a fool, pretend they see the excellent patterns and beautiful colours of the new garments. When the Emperor goes out to show the new robe to his citizens, it is only an innocent child who bursts out: "But he hasn't got anything on" – and even then, the Emperor proudly continues his procession (Anderson, 1837).

Both **Raja, Baggs, Chemero, and Anderson (Raja et al.)** and **Hesp** ask what part of the folktale we are alluding to with the title of our paper. Raja et al. had expected us to "follow the child's

lead" and expose Friston blankets as indeed being nothing more than the Emperor's New Clothes, while Hesp asks us to clarify whether we intended to attribute intentional deceit to free energy principle (FEP) researchers. Let us start by clarifying our position here, which will provide the leitmotif for the rest of our response. First of all, we would like to be clear that our criticism is levelled against the content of academic proposals and the way certain concepts are used, not against the people themselves. To stay with the metaphor, our target article focused on the putative patterns and colours of the clothes, not on the Emperor and his courtiers. Second, there is an important distinction between intentional deceit and what Frankfurt (2005) refers to with the technical term *bullshitting*. Deceit requires knowing the truth and intentionally hiding it from others, while bullshitting involves attempting to persuade without taking truth into consideration. Importantly, bullshitting can be done unintentionally. One way in which bullshitting might happen unintentionally is when speculations are underconstrained, as is sometimes the case when mathematical constructs are applied outside of their original formal context. In such scenarios, the argumentative structure resembles that of ordinary science, but fails to engage meaningfully with the underlying formal constructs. The result is armchair philosophy dressed up as an empirical research programme, which **Allen** goes so far as to label as "cargo cult" science.

Third, the FEP research programme has been an ambitious, daring, and speculative endeavour from the start. This is what attracted each of the authors of the target article to work on the FEP in the first place: the potential for a formalisation of big philosophical concepts like agency, intentionality, life, and consciousness. But in order to make this work, it is necessary to intricately weave together mathematical modelling, philosophy of science, philosophy of mind, and empirical research. This interdisciplinarity, along with the rapid pace at which papers on the FEP are published, introduces vulnerabilities: it is a difficult literature to keep up with, and even experts can sometimes be confronted with things that they do not fully understand. It was this feeling of epistemic dissonance, central to the Emperor's New Clothes, that we wanted to allude to with our title. Andersen's folktale captures a common sentiment among those getting interested in the FEP: am I too incompetent to understand this, are the conclusions that are being drawn not supported by the theory?

Our aim for the target article was to critically assess the current state of both philosophical and scientific literature using Markov blankets within the FEP: what are the assumptions required to transform Markov blankets from a technical notion in statistics and probability theory to an ontological notion used to settle philosophical and scientific disputes? In other words, can we at least agree on what kind of clothes the Emperor is supposed to be wearing?

In analysing the literature, we found a number of technical slippages and conceptual unclarities: an ambiguity between Markov (Pearl) blankets as modelling tools and Markov (Friston) blankets as real-world boundaries, an ambiguity between instrumentalist, realist, and literalist statements about the latter Friston blanket construct, and a conflation of different types of explanatory project that might use either kind of blanket. We hoped that the range of options we presented in the target article could serve as an open invitation for those working with the FEP to make clear what their commitments really are, and how those commitments can support the claims they make.

We are extremely grateful that this challenge has been taken up by a great number of commentators, from both inside and outside the literature on the FEP. We would like to thank the authors of all 35 commentaries for their insightful critiques and friendly suggestions. The commentaries cover a wide range of topics and take many different stances: some push back against our framing of the problems surrounding Markov blankets and the FEP, some suggest ways to fix the problems we identified, and yet others highlight a number of additional problems. In Table R1 we present a thematic overview of the commentaries as we understood them.

The range of attitudes adopted in the commentaries shows just how widespread disagreement and confusion is in the FEP literature about the conceptual, metaphysical, and methodological commitments implied by the use of Markov blankets. Crucially, the issues we raised in the target article are interrelated: you can only claim that the structures identified in your model carve out real boundaries in the world if the mapping between model and target is structure preserving (see **Parr**). The usefulness of instrumentally treating organisms *as if* they are models of their environments does not imply that it will also be useful to treat them *as if* their physical boundaries emerge within this model. The proven utility of Markov blankets as modelling tools in the probabilistic inference literature does not directly support the claim that to *be* a thing is to *have* a Markov blanket. There is a thin line between a scientific endeavour whose metaphors are pushed just a bit too far, and a heavy-duty metaphysical theory dressed up as an empirical research programme. To better capture the diversity of views on offer we have divided our reply into several, mutually supporting sections, with a number of commentaries being mentioned in multiple sections.

R2. Are Friston blankets just Pearl blankets?

The first major issue we would like to tackle in our reply to the commentaries is the distinction (or lack thereof) between Pearl and Friston blankets. One of the major contributions of the target article was to make explicit the previously unacknowledged shift from Pearl blankets to Friston blankets. A number of commentaries argue that the two kinds of entities are, in fact, cut from the same cloth, and that our distinction just captures two different ways of referring to the same Markov blanket formalism.

Parr and **Friston** point out that a Markov [Pearl] blanket is defined in terms of conditional probabilities and, therefore, can only be delineated in the context of a model described in terms of probability distributions. Whether the distributions involved will be sourced from a steady-state density of some dynamical system, or just some static model (i.e., one not evolving over time) should not make a difference for the identification of a Markov [Pearl] blanket. **Ramstead** claims that Friston blankets simply *are* Markov blankets, that is, that Friston blankets and Pearl blankets are the same abstract mathematical objects denoting conditional independence, deployed in different modelling contexts. **Hesp** concurs and points out that much of the confusion we identified stems from differing background assumptions about the kinds of causal relationships and types of systems that Friston and others are trying to study. This is echoed by **Kiverstein and Kirchhoff** who write that "Friston blankets are interpretations given of the Markov blankets formalism in the context of the FEP that purport to describe autopoietic processes." But are Friston blankets and Pearl blankets really one and the same thing?

Our answer is a resounding no, and we also deny that there is any straightforward route to derive Friston blankets from Pearl blankets. The difference between Pearl blankets and Friston blankets is not simply due to differing causal commitments, as

Table R1. Thematic overview of commentaries

Theme	Focus	Commentaries
Formalism	Markov blanket formalism	Virgo, Rosas, and Biehl; Friston; Aguilera and Buckley
	Free energy and physics	Spector and Graham; Stoffregen and Heath; Ramstead
	Spatial boundaries	Parr
Philosophy of science	Realism, instrumentalism, and literalism	Sánchez-Cañizares; Kiefer and Hohwy; Williams; Kiverstein and Kirchhoff; Ramstead; Rorot, Korbak, Litwin, and Miłkowski; Hipólito and van Es; Friston; Colombo; Seth, Korbak, and Tschantz; Wiese; Menary and Gillett
	Causality and interventions	Btesh, Bramley, and Lagnado; Yon and Corlett
	Reification	Andrews
	Unification	Gomez-Marin
	Models and abstraction	Nave; Spiegel; Ciaunica; Beck
Philosophy of mind and life	Internalism and externalism	Facchin; Menary and Gillett
	Autopoiesis	Nave; Suzuki, Miyahara, and Miyazono; Raja, Baggs, Chemero, and Anderson; Kiefer and Hohwy; Dengsø, Robertson, and Constant
	Ecological psychology	Stoffregen and Heath; Raja, Baggs, Chemero, and Anderson
	Agents and selves	Sprevak; Colling; Seth, Korbak and Tschantz; Yon and Corlett; Wiese
Interdisciplinary	Boundary objects	Fox; Allen
	Evolution	Veit and Browning

suggested by **Btesh, Bramley, and Lagnado** (**Btesh et al.**), since the use of Markov blankets (Pearl or Friston) is not tied to models interpreted under some causal semantics (Pearl, 2009). Similarly, commentaries such as **Hesp**, **Kiverstein and Kirchhoff**, and **Wiese** that either invoke, or seem to imply the necessity of, causality in Bayesian models, fall short of explaining where this causal interpretation comes from in the first place. Furthermore, Friston blankets are not just Pearl blankets transformed by technical assumptions (cf. Biehl, Pollock, & Kanai, 2021), or buttressed with additional philosophical commitments. In his commentary, **Friston** correctly argues that Markov (i.e., Pearl) blankets are merely statements about conditional independence, such that for generic Markov chains "the 'present' is a Markov [Pearl] blanket that separates the 'past' from the 'future'." However, this is apparently not what a Friston blanket is, as Friston himself (Friston, Da Costa, & Parr, 2021a) writes:

> "For Markovian systems, the states at the current time are the blanket states that separate states in the future from states in the past. However, these are not Markov [Pearl] blankets of the steady-state density."

This tells us that the past and future of a random variable, which would be part of a "naive" Pearl blanket in dynamic settings (Pearl, Geiger, & Verma, 1989), are not part of a Friston blanket. To really get to the core definition of what a Friston blanket is, we can further look at a reply to the example found in Biehl et al. (2021) that describes their mathematical assumptions. Friston et al. (2021a) once again helpfully explains that

> "[…]; the FEP only applies to Markov blankets that emerge under sparse flows; in particular, when autonomous states are uncoupled from external states (by definition)."

Here, "sparse flows" refers to a specific coupling structure required between partitions (internal, external, active, and sensory states) in dynamical (i.e., time-evolving) settings (see equation [7] in Friston et al. [2021a] or equation [12] in Friston et al. [2022]). More recently this also goes by the name of the "sparse coupling conjecture" (Friston, Heins, Ueltzhöffer, Da Costa, & Parr, 2021b; Friston et al., 2022-version 1, removed in version 2). In practice, this has so far been presented as an arbitrary assumption, a conjecture, that excludes examples such as the one given by Biehl et al. (2021) from falling within the scope of the FEP. Furthermore, there are multiple ways to generalise conditional independence relations to dynamic Bayesian networks,[1] which ultimately means that Friston blankets are not just *the* natural, or even unique, description of conditional independence applied to dynamical systems. Contrary to **Ramstead**'s claims, there are reasons to distinguish Pearl and Friston blankets: a Pearl blanket is just a statement about conditional independence, while a Friston blanket is a finely crafted posit that includes both conditional independence and a number of non-trivial additional assumptions that are necessary in order for the construct to play a particular role in the wider FEP theory.

As we have already seen, the debate over the distinction between Pearl blankets and Friston blankets reveals cracks in the conceptual foundations of the wider free energy framework, with past co-authors disagreeing among each other about the legitimacy of this distinction. **Aguilera and Buckley** and **Virgo, Rosas, and Biehl** (**Virgo et al.**) both speak in favour of the distinction. As these authors point out, finding Markov blankets that can delineate boundaries or sensorimotor loops within the bounds of the assumptions made by the FEP is far more difficult than commonly thought. Quite tellingly, **Friston** himself seems to be conflicted about the difference between the two kinds of blankets. On the one hand, he writes that "Pearl and Friston blankets are just Markov blankets in the usual Markovian sense (Pearl, 2009)." On the other, he follows this directly with a rhetorical question that undermines his previous claim: "Are Markov blankets used in an ontological sense under the free energy principle (FEP)? Yes." This statement not only undermines the previous one, as differentiating between the purely technical and more

metaphysically laden uses of the Markov blanket formalism was the whole point of the target article, but also undercuts **Andrews**' accusation that we somehow misread or misinterpret the free energy theorists' metaphysical intentions.

A similar kind of confusion can be found in **Fox**'s proposal that the distinction between the two kinds of blankets will only lead to further debate and will not facilitate modelling complex systems. Instead, he advocates for the position that "it may be useful to frame systems in terms of constructs such as Markov blankets, but without applying all technical details and associated mathematics." We fail to see how this is a more constructive solution or how it could help to alleviate, rather than exacerbate, the problems Fox attributes to our account. After all, a lack of clear distinction between formal technicalities and ontological commitments is what got the field into the situation it is in now. It is to the latter topic that we turn next, namely the metaphysics and ontology implied by the use of Friston blankets to demarcate worldly boundaries.

R3. The metaphysics and ontology of Friston blankets

Several of the commentaries on our target article focused on the metaphysical and ontological aspects of our critique, namely whether the use of Friston blankets to demarcate real-world boundaries requires a commitment to either *literalism* or *realism* about these constructs, and also what exactly (scientific) realism about theoretical entities of this kind really commits one to. One initial point we want to pick up on is the importance of distinguishing between *local* scientific realism (i.e., about our attitude towards some particular theoretical entity, such as Friston blankets) and *global* scientific realism (i.e., about our attitude towards theoretical entities in general). It is possible to be a global scientific realist while still advocating anti-realism or instrumentalism towards some particular theoretical construct, which is the position we took ourselves to be suggesting towards the end of the target article. We did not, as **Kiverstein and Kirchhoff** suggest, intend to argue against scientific realism in general, or even to enthusiastically endorse instrumentalism about Friston blankets, as both **Sánchez-Cañizares** and **Colombo** took us to be doing (rather, we think that instrumentalism might be the best option if one is really committed to making use of Friston blankets, instead of simply discarding the construct as unworkable). Other commentators, such as **Andrews**, **Hipólito and van Es**, and **Ramstead**, do seem to be advocating for a general anti-realism or instrumentalism, which would render the particular question of the status of Friston blankets somewhat less pressing (if all theoretical entities are just instrumental tools, then the ontological status of any particular theoretical entity does not matter so much, or perhaps at all).
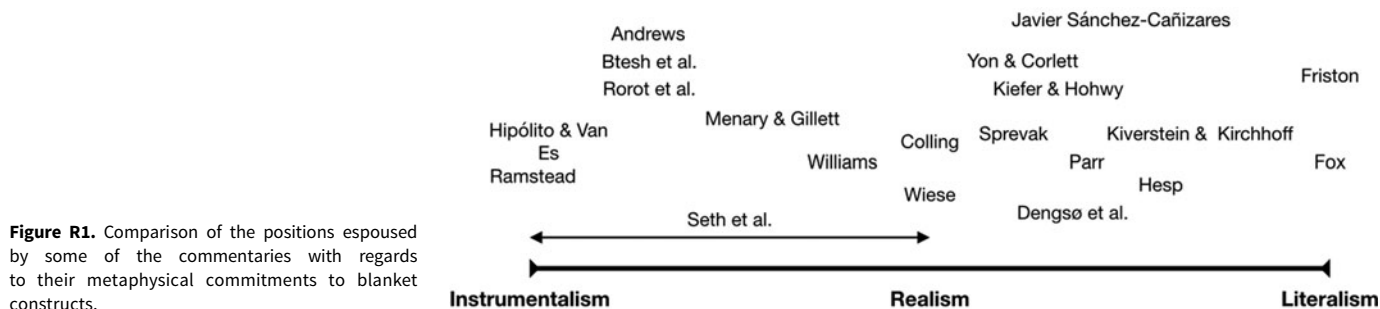
This brings us to the next point we would like to comment on, which is that we were quite struck by the sheer diversity of ontological attitudes exhibited by proponents (or at least defenders) of both the FEP, and the role of Markov blankets within it. These range from a steadfast realism of some variety, as in the cases of **Kiefer and Hohwy, Kiverstein and Kirchhoff**, and **Wiese**, the more cautious realism exhibited by **Sprevak** and **Seth, Korbak, and Tschantz** (Seth et al.), all the way to the full-blown instrumentalism of **Andrews**, **Hipólito and van Es**, and **Ramstead**. At the very least we hope that our target article, and the commentaries it elicited, have demonstrated that the ontological status of Markov blankets within the FEP is far from settled, and that those working within the framework might do well to

better communicate their ontological commitments, both to each other and to the outside world.

In his commentary on our target article **Friston** claims that Markov blankets "are deployed in various scientific fields to 'carve nature at its joints'," while nonetheless conceding that "there are many ways of carving nature at its joints," which seems to us to be at best hedging the ontological status of Friston blankets. He also expresses yet another ontological option, which is the view that Friston blankets are literally real *because* the real world itself is at base composed of mathematical structures of some kind. This is certainly not scientific realism in the traditional sense, but rather something closer to the "it-from-bit" hypothesis that the universe is fundamentally computational or informational, defended for example by Wheeler (1982), Zuse (1982), and Wolfram (2002). Menary and Gillett (2020) have already indicated that there are many unanswered questions about using this kind of formal ontology to ground Markov blankets, and in their commentary (**Menary & Gillett**) they raise further insightful issues with Friston's claim that Markov blankets could ever be used to carve nature at its joints. **Colombo** also suggests interpreting the literalist ontology in this way, and doubts whether a purely instrumentalist interpretation of Friston blankets is either as successful or uninteresting as we made it out to be.

What this points to is the distinct lack of clarity surrounding Friston's own ontological commitments, which appear to range from fully instrumentalist (e.g., Ramstead, Friston, & Hipólito, 2020) to fully realist (e.g., Friston, Wiese, & Hobson [2020], and in his response to our target article). Insofar as Friston's work on Markov blankets in the FEP is the canonical starting point for much of the research that we discussed in the target article, any appeal to the "standard" use of Markov blankets in the framework is going to be necessarily vague. This is not to say that everyone working on the FEP must agree about their ontological commitments (far from it, why not let a thousand flowers bloom), but rather that researchers must be careful to make their own commitments clear and not to assume that this is a settled question on which they can simply adopt the party line.

Several commentators suggested that our three-fold distinction between literalism, realism, and instrumentalism might either be too strict or unclear in some respects – we freely admit to both charges, and we are happy to see the conversation that we started being carried forward in a positive direction. Of particular note here is **Seth et al.**, who proposed that it might be better to think in terms of a spectrum of options ranging from literalism to instrumentalism, with various forms of realism lying in between. We are sympathetic to this suggestion (see Fig. R1), but we disagree with their categorisation of Pearl blankets at the instrumentalist end of this spectrum and Friston blankets at the realist end. Pearl blankets, as we understand them, are indeed just a statistical modelling tool, but such a tool could be used to identify "real" features of the world (just probably not ones that are directly equivalent to the blankets themselves); and while Friston blankets are intended as a more metaphysically robust entity, one could treat this entity in an instrumentalist fashion, as several of the other commentaries demonstrate. The distinction between Pearl blankets and Friston blankets does not directly correspond to any particular position along the realism–instrumentalism spectrum; one can be a global instrumentalist and make use of Friston blankets, or a global realist and make use of Pearl blankets. A similar claim is made by **Williams**, who understands our distinction "as delineating between the formal and the physical." Williams expands on this by

Figure R1. Comparison of the positions espoused by some of the commentaries with regards to their metaphysical commitments to blanket constructs.

proposing that the two kinds of blankets are best understood as being located on different levels of Marr's computational analysis. However, we think that viewing our distinction through the lens of Marr's heuristic division can only blur the boundaries between the instrumental and realist uses of the Friston blanket construct. The fact that a problem solved by a target system can be described using a Markovian formalism does not require a commitment to Markov blankets being explicitly represented (or instantiated) by the system itself. It is only once theorists commit to the target system explicitly possessing or instantiating Markov blankets, either as representations or boundaries, that we can talk about realism.

Some commentators also questioned whether our analysis of the metaphysical uses and misuses of Markov blankets really hits the mark. **Kiverstein and Kirchhoff** argue that our critique relies on what they call the "literalist fallacy," the assumption that realism about the boundaries picked out by Friston blankets implies a belief in the literal existence of Markov blankets in the physical world. This latter belief is indeed what we meant by "literalism," and at least some proponents of the FEP seem to be committed to it. It is hard to see what else is meant when, for example, Allen and Friston (2018) write "the very existence of a system depends upon conserving its boundary, *known technically as a Markov blanket*" (emphasis added) – even if the authors intended to say something else, the most natural reading here is a literalist one, where the physical boundary of a system is identified as a Markov blanket. On the other hand, we by no means wanted to deny the possibility of a more moderate "realism" about the kinds of boundaries that might be picked out by Friston blankets. However, even this more moderate view is not without issues of its own, which we discussed extensively in the target article, and were also mentioned in some of the commentaries. These issues are independent of a literalist interpretation of Friston blankets, and therefore cannot be dismissed by appealing to a "literalist fallacy." At the very least, realism about Friston blankets requires a formalism that guarantees a stable and non-arbitrary mapping between formal model and target system. As it currently stands, the proposed formalism has only been applied to toy models in which such a mapping is already taken for granted, and we have identified in our target article a number of issues that make a stable and non-arbitrary mapping highly unlikely in more realistic (and thus more interesting) cases (see also Aguilera, Millidge, Tschantz, & Buckley, 2021).

Meanwhile, **Andrews** is concerned that while we accuse some proponents of FEP of *reifying* the Markov blanket construct, we fail to give a clear definition of reification and risk conflating it with the metaphysically innocent employment of idealised models. We are happy to at least provisionally accept Andrews' proposed definition of reification as "the mismapping of formal structure onto target phenomena – or theoretical representation

thereof – in a manner that leads us to misapprehend the causal structure of nature," although we might add that it is not just the *causal* structure that is at stake here, but rather the structure of the real world more generally (whether or not that structure is either primarily or wholly causal is a further question that we will not attempt to answer here). With this definition on the table, Andrews then argues that our accusations of reification miss the mark, as the use of Markov blankets in the FEP is in fact a "mere modelling gambit" that does not aim to directly uncover the causal structure of the world. Furthermore, Andrews claims that the FEP (or variational free energy minimization) framework does not even *in general* aim to "generate knowledge of natural systems," but is rather just a modelling framework that must be coupled with further auxiliary assumptions before it can generate empirically testable hypotheses. This approach might be congruent with what *some* FEP theorists are doing, and we applaud them for exercising epistemic caution with their uses of the framework, but it seems clear to us that others have taken it far more literally than this, especially when it comes to the use of Friston blankets to demarcate worldly boundaries. Andrews also suggests that our own claim that Markov blankets are "substantiated by the empirical literature" is a case of reification, as formal tools of this kind are not the sort of thing that can be empirically substantiated. Our claim here was not that Markov blankets are empirically substantiated in the sense of being proven to exist, but rather that their repeated application in the empirical literature vindicates their usefulness as a modelling tool – but we are happy to acknowledge that we may have misspoken when making this point, and would invite those whom we have accused of reification to do the same, or else to explain more precisely what they think Friston blankets can tell us about the world.

Finally, some commentaries (**Beck**; **Ciaunica**; and **Dengsø, Robertson, & Constant [Dengsø et al.]**) discuss the metaphysical possibilities afforded by Pearl blankets and Friston blankets, focusing on the question of how to legitimately connect the map with the territory. Ciaunica emphasises the role of the *cartographer* (i.e., the modeller) in this process, connecting this to older debates between Plato and Aristotle, and somewhat more recent debates between Schlick and Neurath. Beck similarly emphasises the importance of the domain of application of a model, but ultimately considers "metaphysics a pointless endeavor," since "all we have are models." While we agree that it is important to ask who is using a model and for what purpose, we also think that it is important to make clear any theoretical presuppositions about how models connect to the world, that is, what metaphysical or ontological attitude one is taking towards the claims made about Markov blankets and the FEP. Dengsø et al. argue that in order to account for the dynamics of living systems (and to

avoid what they call preformationism), Markov blanket theorists need to adopt a process ontology. This is an intriguing option, furthermore demonstrating that many of the interesting philosophical moves available to proponents of FEP are not intrinsic to the framework itself, but rather involve additional metaphysical presuppositions that are necessary in order to make the use of Friston blankets plausible and coherent. In the next section we will accordingly consider what work, if any, Friston blankets *alone* are able to do.

R4. What Friston blankets are not doing

Having defended the validity of our distinction between Pearl and Friston blankets, it is finally time to consider what kind of work Friston blankets, at least in their present form, actually can, and more crucially *cannot*, do. For instance, the causal language adopted by several commentaries reveals the confusion we find when inference with a model is treated as equivalent to inference within a model (as explicitly done by **Kiverstein & Kirchhoff**, **Hesp**, and **Wiese**, and implicitly suggested by **Friston** and **Parr**). This confusion stems from using dependencies intrinsic to models, for example, the fact that changing variable *A* has an effect on variable *B*, to make causal claims about the world extrinsic to that model. We can unpack this better by appealing to an example from Beer's (2004, 2014, 2020) studies of stable patterns in the Game of Life. Do the five rules constituting the game describe interactions that can be treated as causal within the Game of Life universe? Yes, they do. Do the same rules count as causal in our own universe? No, they don't, since the structures described by the Game of Life cannot be straightforwardly mapped onto the structures of our universe. That being said, we could find applications of the Game of Life where it would make sense to treat it as a model of causal relationships between entities in the real world. However, this would only mean that the game has been given a causal interpretation that can also be applied to the dynamics of our world, not that it straightforwardly captures what causation is, or that the Game of Life can by itself be used to discover anything about our world.

Looking closer at the technical commitments defining Friston blankets, the steady-state assumption behind the FEP is perhaps the most controversial one. This assumption naturally leads to questions regarding the relevance of evolutionary dynamics and history more generally (**Veit & Browning**), given the asymptotic independence on initial conditions implied by the even stronger conditions of ergodicity and weak-mixing often assumed by the FEP (see also Di Paolo, Thompson, & Beer [2022] for a related discussion, and Da Costa, Friston, Heins, & Pavliotis [2021] for a technical treatment without assuming ergodicity and weak-mixing, but retaining stationarity). The detailed commentary by **Virgo et al.** demonstrates that a definition of blankets based on steady-state distributions ultimately leads to systems without memory. As clearly portrayed in their example, this is because of the data processing inequality (Cover & Thomas, 2006) applied to Friston blankets, explaining how the internal states of a system of interest (perhaps an agent) cannot store any more information than is present in their current observations at any one point in time. This leads to a severe implication that systems under Friston blankets can only do inference on information located in the perceptual present – past experiences carry no weight for such systems. The only way to allow for some form of memory is to essentially break the stationarity assumption, at least temporarily, as seen in Parr, Da Costa, Heins, Ramstead, and Friston

(2021). The implication of breaking this assumption is that the FEP is actually doing its most interesting work *away* from steady state, and therefore *without* Friston blankets. This seems to contradict the large body of work on models based on active inference formulations where past information appears to play a crucial role (see for instance Baltieri & Isomura, 2021; Friston et al., 2015, 2017a, 2017b; Isomura, Shimazaki, & Friston, 2022; Lanillos et al., 2021; Mazzaglia, Verbelen, Çatal, & Dhoedt, 2022; Parr & Friston, 2019 – to give only a few recent examples). It is crucial, however, to highlight that the stipulative nature of Friston blankets is not playing any role in these models – active inference can and does exist without Friston blankets.

The so-called "sparse coupling" assumption, a conjecture that essentially allows the definition of Friston blankets for dynamics starting from Pearl blankets on a steady-state distribution, also leads to questionable consequences. **Aguilera and Buckley** analyse the validity of this assumption, arguing that its high specificity makes it unlikely to realistically characterise relevant (natural and physical) boundaries. In discussing a different derivation of Friston blankets, namely an asymptotic approximation to a weak-coupling equilibrium (Friston et al., 2021b), they also analyse how sparse coupling, however unlikely, is necessary to avoid more dramatic shortcomings: conditional independencies that cannot be guaranteed because of time-dependent relations across different components of a system (cf. **Virgo et al.**). Altogether, we have an assumption that appears to be essential to Friston blankets, sparse coupling, that may unfortunately just not be able to capture any interesting properties of sensorimotor loops, agents, or living systems.

Moving beyond these somewhat technical details, we believe that there are even more fundamental grounds to question the role of Friston blankets within and beyond the FEP. The main worry is captured well by claims that Markov blankets (whether Pearl or Friston) are "formal tools […that] 'carve nature at its joints'" (**Friston**). In the target article, we present and discuss different examples that show why this is simply not the case. In the primordial soup model (Friston, 2013), the experimenter chooses which set of states to designate as internal and which as external, essentially determining *a priori* where the blanket ought to be. Similarly, in the patellar reflex example, which should be properly and more correctly understood following **Spiegel**'s proposed alternative setup, the experimenter once again chooses what constitutes the set of internal states from the vantage point of a scientist who could be asking different possible questions. Contrary to what **Hesp** suggests, our goal with this example was to highlight the ambiguous role of co-parents within a single time slice (as required by Friston blankets), not to look at dependencies across time steps (which are not a part of Friston blankets anyway, as explained above following Friston et al., 2021a). Friston seems to imply that this aspect of experimenter (or modeller) choice is a *feature* of his formulation, but as **Suzuki, Miyahara, and Miyazono** (**Suzuki et al.**) point out, this sidesteps the real question of whether Friston blankets are patterns already present in the "natural" cloth, or patterns that are intentionally introduced by an external agent (either by the tailors themselves, or perhaps just by the imagination of the courtiers).

In a series of commentaries addressing this very issue, it seems to be widely recognised, and in some cases welcomed (**Andrews**; **Hipólito & van Es**; **Kiverstein & Kirchhoff**; **Sprevak**; **Wiese**), that Friston blankets are not foundational to any theory of systems, things, sensorimotor loops, or agents. The heavy lifting is done by a scientist injecting their own assumptions into a

model (**Facchin**; **Hipólito & van Es**; **Kiverstein & Kirchhoff**; **Menary & Gillett**; **Raja et al.**; **Rorot, Korbak, Litwin, & Miłkowski** [**Rorot et al.**]; **Sprevak**; **Suzuki et al.**; **Wiese**). For example, the choice of which parts (particles) should count as internal in the primordial soup model is essentially arbitrary. As has already been stated, **Friston** believes this is a deliberate and advantageous feature. Some of his defenders view this simply as part of standard modelling practice in science (**Andrews**; **Hipólito & van Es**; **Kiverstein & Kirchhoff**; **Sprevak**; **Wiese**), and not a reason to dismiss the use of Friston blankets without empirical tests validating their practical applications. However, others (**Facchin**; **Menary & Gillett**; **Raja et al.**; **Rorot et al.**; **Suzuki et al.**) agree with the critical position we laid out in the target article, and call for clarity in how the explanatorily relevant blankets are identified, and who or what is doing this identification. These two positions are clearly not inconsistent, but they do require that we carefully consider any explanatory role attributed to Friston blankets. Friston blankets cannot be instrumental in defining interesting boundaries, if such boundaries have to be deliberately selected ahead of time for this tool to be successfully applied to. As eloquently argued by **Colling**, a "formalism itself does not licence predictions about which systems are amenable to the formalism and which are [not]." Colling argues that this problem has been faced by other frameworks in the past, such as dynamical systems theory, and points out that defenders of Friston blankets need to "provide an explanation or prediction of which systems are amenable to the formalism – or because the formalism is applicable to every 'thing' (Friston, 2019), which systems are amenable to specific applications of the formalism independent of the particular application of the formalism itself."

This then leads us to some of the core debates regarding teleology, autopoiesis, and the foundational principles of enactive cognitive science in relation to Friston blankets and the FEP, which were discussed in several commentaries. **Veit and Browning** argue that ignoring the teleonomic context of biological systems is ultimately what is causing a lot of the confusion surrounding Friston blankets since previous research has largely overlooked questions about the evolutionary history or function of biological boundaries. We sympathise with this concern, but we worry that the confusion surrounding these issues will only grow, as the ambitions of FEP's proponents seem to have recently moved away from the idea of seeking to explain *only* biological systems (Friston, 2010; Friston, Kilner, & Harrison, 2006) to explaining systems in general (Da Costa et al., 2021; Friston, 2019; Friston et al., 2022). This move could call into question the necessity of teleonomics, but at the same time also raises the issue of whether the FEP is still even trying to say something specific about biology, or cognition, at all. **Raja et al.** and **Suzuki et al.** (see also Di Paolo et al., 2022) carefully observe that Friston blankets are not intrinsically self-determined properties of a system, because as we said above they require a scientist to specify an internal partition, and thus fall short of capturing the *autopoietic* mechanisms found in natural systems (contrary to what, e.g., **Hesp**, **Kiverstein & Kirchhoff**, and **Seth et al.** claim). **Nave** argues forcefully that FEP's abstracting away from the "metabolic turnover" of living systems leads one to "fundamentally misconceive what an organism is." Simply put, FEP ignores that a biological agent's structure not only constrains its own dynamics, but that the dynamics also constitute its structure (cf. Jonas, 1966).

So, what can Friston blankets actually do? It seems clear, at this point, that a Friston blanket-centric metaphysics that makes claims about "blankets of the mind" or "blankets of life" is deeply problematic. It is most certainly not a project that can be accomplished by "just doing the maths," since the relation between Friston blankets and Pearl blankets is all but straightforward, and Friston blankets currently appear to be afflicted by some potentially disqualifying limitations. Furthermore, it should not be presented as a principles-first approach for defining which entities do inference within a model (Friston, 2019), as such entities have to be specified in advance by the modeller. Some authors, such as **Wiese**, acknowledge that the framework is primarily interested in defining a fundamental metaphysics, rather than realising a purely empirical research programme, and we think that this approach at least offers a more perspicuous starting point for future research.

R5. Not everything needs to be a blanket

Having defended the distinction between Pearl blankets and Friston blankets, discussed leading approaches to the metaphysical commitments of the latter, and explored the kind of explanatory work they might be used for, we now conclude our response to the commentaries by looking ahead to the future of Friston blankets and the FEP in science and philosophy.

The most obvious way forward has already been proposed in our target article and involves the tedious, but manageable, task of cleaning up and clarifying the use of the formal constructs in the FEP literature. While we do not expect to see a monolithic consolidation of the framework, confusion about Friston blankets being "just" Pearl blankets still pervades the majority of work on the FEP and must be addressed somehow. There are different interpretations as to who tailors the blankets (**Facchin**; **Friston**; **Nave**; **Raja et al.**; **Suzuki et al.**), what their role is (**Aguilera & Buckley**; **Menary & Gillett**; **Parr**; **Sprevak**; **Wiese**), and how we should even define a blanket (**Btesh et al.**; **Kiverstein & Kirchhoff**; **Virgo et al.**). The only point that remains uncontroversial is that in order to try and make use of the FEP and its associated formal tools, authors ought to explicitly state what their starting assumptions are.

A viable, but perhaps less appreciated, alternative is to simply make use of the FEP without committing to Friston blankets. While the literature on the FEP has recently been centred around the definition and use of Friston blankets, this hasn't always been the case. As the generalisation of conditional independence relationships to dynamical settings is not in principle unique to Friston blankets, one could simply abandon Friston blankets for a different construct. This could avoid some of the construct's current shortcomings (as discussed by **Aguilera & Buckley** and **Virgo et al.**) by introducing more transparent and explicit assumptions and getting rid of the unhelpful "just do the maths" rhetoric. More interestingly, however, broader work on active inference, predictive coding, and prediction error minimization has been flourishing mostly independently from the notion of Friston blankets (see, e.g., Lanillos et al., 2021; Mazzaglia et al., 2022; Millidge, Seth, & Buckley, 2021; Spratling, 2016 for reviews in neuroscience, robotics, and machine learning). While the mathematical connections between these ideas are unquestionable (especially under Gaussian assumptions), their level of commitment to the full tenets of the FEP varies greatly. This showcases that active inference by itself could be used as a simple alternative to reinforcement learning formulations of behaviour and decision making, one that appeals to a (different) cost function (Millidge, Tschantz, Seth, & Buckley, 2020), without committing to the

FEP's extra baggage. Importantly, a version of the FEP that does not appeal to Friston blankets need not collapse into an inherently instrumentalist position, detached from any semblance of empirical validation (about which we are warned by **Colombo**). An active inference alternative to reinforcement learning should be judged on its own merits, according to the empirical results that it generates and the theoretical virtues that it demonstrates.

A third, and yet more radical, answer to the issue of FEP's future brings our attention to the framework as a whole. Although we have already provided a set of actionable solutions to fix the problems surrounding Friston blankets, and have shown that many of the models from the FEP toolkit have seen continuous development outside of the framework, we would like to end by asking about the overall benefit of the FEP for science. With **Allen**'s accusation of a "cargo cult" and **Gomez-Marin**'s worries that the FEP might be "not even wrong" and "appears sufficiently vague to be immune to empirical data," it is worth considering that we might just not need the FEP at all.

Nonetheless, if one wishes to adopt this radical strategy, it is important to do it for the right reasons. For instance, **Spector and Graham** argue that quantities defined by the FEP are needlessly obfuscating the differences between information-theoretic and thermodynamical meanings of terms like "energy," and cannot be easily tied to a meaningful physical interpretation. However, this simply follows a trend in modern stochastic thermodynamics (where the focus is on far from equilibrium systems), and is not as unique or devastating to the FEP as the authors make it out to be. Unlike the classical equilibrium case discussed by these authors, thermodynamic quantities such as heat, free energy, and so on are usually not well defined far from equilibrium, and are best tackled by a formulation based on information theory/geometry (for a recent review see for instance Kim, 2021). Further discussion on the relation of the FEP to physics can be found in Friston (2019) and Friston et al. (2022), which now clarify that the free energy invoked by the FEP is not the same as the free energy defined in thermodynamics.

Similarly, both **Raja et al.** and **Stoffregen and Heath** take the FEP to be incongruent with the commitments of ecological psychology, and see this as a reason to discard it. Stoffregen and Heath seem overly impressed by talk of "inference" within FEP, which does not need to have the intellectualistic commitments that they think it does (cf. Bruineberg & Rietveld, 2014). Raja et al. state that the FEP cannot model relational properties, such as affordances. However, they disregard a productive line of research within ecological psychology that does seek integration with selectionist neuroscience, of which FEP is a variety (see, e.g., Bruineberg & Rietveld, 2019; Reed, 1989, 1996).

In searching for alternatives to the FEP, it might be useful to follow **Btesh et al.** and **Yon and Corlett**, and consider whether a more intervention-focused causal framework (in the sense of Pearl, 2009) could lead to more insightful explanations of cognition. As **Btesh et al.** argue, taking inspiration from an established literature on causality in animal cognition, non-causal alternative frameworks have so far fallen short of explaining any truly high-level features of intelligent and cognitive systems (Marcus, 2018; Pearl & Mackenzie, 2018). On the other hand, Yon and Corlett propose that while Friston blankets may not be the right solution, a different type of blanket, which they named a "cognitive blanket" (cf. the cognitive domain in autopoiesis [Beer, 2014]), based on causal interventions à la Pearl, may provide a more direct bridge to

inferential theories of cognition where brains act as hypothesis testing machines by intervening on the environment. Such a framework should still be mindful of the fundamental dichotomy drawn in our target article, between inference *with* or *within* a model.

Overall, we think that a plurality of methods is likely to be the most fruitful approach for a cohesive study of all the diverse aspects of cognition and mind, as evidenced by the wide range of perspectives offered in the 35 commentaries that we received. The tools offered by the FEP, including the new construct of a Friston blanket, might, with some refinement, become a valuable addition to this plurality of methods, but we should keep in mind that not every boundary needs to be a blanket.

R6. Conclusion

In the target article we critically assessed the current state of the philosophical and scientific literature using Markov blankets within the FEP, finding a number of technical slippages and conceptual unclarities. We hoped our target article could serve as an open invitation for those defending the FEP to make clear what their commitments really are, and how those commitments can support the claims they make about Markov blankets. The 35 commentaries we received show that there is widespread disagreement about the role of the Markov blanket construct within the FEP, its ontological status, and the explanatory projects in which it is embedded. Therefore, we do think that the Emperor is currently naked, or at least wearing a different costume each time that he appears. Some commentaries provided fruitful suggestions for how best to redress the Emperor – time will tell whether these attempts will be successful. We hope that the target article, the commentaries, and our reply will all contribute to a more nuanced and productive discussion moving forward.

Note

1. Some alternatives to Friston's preferred steady-state assumption + sparse coupling conjecture include:

(a) Friston's very own alternative, that is, an asymptotic approximation to a weak-coupling equilibrium, found in Friston et al. (2021c) and discussed by **Aguilera and Buckley**,

(b) Biehl's example, dropping the above conjecture while retaining the steady-state assumption and adding other constraints, as in Appendix A in Biehl et al. (2021) and Appendices A and B in Friston et al. (2021a),

(c) definitions that don't require stationarity or even Markov properties, based on computational mechanics (Rosas, Mediano, Biehl, Chandaria, & Polani, 2020),

(d) postulating blankets which are not relegated to a single time slice and thus that can in principle consider history (**Virgo, Rosas, & Biehl**), or just

(e) one of the existing alternatives before the advent of Friston blankets, for example, Flesch and Lucas (2007), Materassi and Salapaka (2014), or Koster (1999) in particular for something related to the worries about directed cyclic graphs raised by **Aguilera and Buckley**.

References

Aguilera, M., Millidge, B., Tschantz, A., & Buckley, C. L. (2021). How particular is the physics of the free energy principle? *Physics of Life Reviews*.

Allen, M., & Friston, K. (2018). From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese*, 195(6), 2459–2482.

Anderson, H. C. (1837). The Emperor's New Clothes. English translation by Jean Hersholt available at https://andersen.sdu.dk/vaerk/hersholt/TheEmperorsNewClothes_e.html

Baltieri, M., & Isomura, T. (2021). Kalman filters as the steady-state solution of gradient descent on variational free energy. *arXiv preprint*, arXiv:2111.10530.

Beer, R. D. (2004). Autopoiesis and cognition in the game of life. *Artificial Life*, *10*(3), 309–326.

Beer, R. D. (2014). The cognitive domain of a glider in the game of life. *Artificial Life*, *20*(2), 183–206.

Beer, R. D. (2020). An investigation into the origin of autopoiesis. *Artificial Life*, *26*(1), 5–22.

Biehl, M., Pollock, F. A., & Kanai, R. (2021). A technical critique of some parts of the free energy principle. *Entropy*, *23*(3), 293.

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, *8*, 599. https://doi.org/10.3389/fnhum.2014.00599

Bruineberg, J., & Rietveld, E. (2019). What's inside your head once you've figured out what your head's inside of. *Ecological Psychology*, *31*(3), 198–217.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Wiley.

Da Costa, L., Friston, K., Heins, C., & Pavliotis, G. A. (2021). Bayesian mechanics for stationary processes. *Proceedings of the Royal Society A*, *477*(2256), 20210518.

Di Paolo, E., Thompson, E., & Beer, R. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 1–39.

Flesch, I., & Lucas, P. (2007). Independence decomposition in dynamic Bayesian networks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty* (pp. 560–571). Springer, Berlin, Heidelberg.

Frankfurt, F. (2005). *On bullshit*. Princeton University Press.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

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

Friston, K. (2019). A free energy principle for a particular physics. *arXiv preprint*, arXiv:1906.10184.

Friston, K., Da Costa, L., Sajid, N., Heins, C., Ueltzhöffer, K., Pavliotis, G. A., & Parr, T. (2022). The free energy principle made simpler but not too simple. *arXiv preprint*, arXiv:2201.06387.

Friston, K., Heins, C., Ueltzhöffer, K., Da Costa, L., & Parr, T. (2021b). Stochastic chaos and Markov blankets. *Entropy*, *23*(9), 1220.

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

Friston, K. J., Da Costa, L., & Parr, T. (2021a). Some interesting observations on the free energy principle. *Entropy*, *23*, 1076.

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

Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017a). Active inference: A process theory. *Neural Computation*, *29*(1), 1–49.

Friston, K. J., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, *100*(1), 70–87.

Friston, K. J., Parr, T., & de Vries, B. (2017b). The graphical brain: Belief propagation and active inference. *Network Neuroscience*, *1*(4), 381–414.

Friston, K. J., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, *6*(4), 187–214.

Isomura, T., Shimazaki, H., & Friston, K. J. (2022). Canonical neural networks perform active inference. *Communications Biology*, *5*(1), 1–15.

Jonas, H. (1966). *The phenomenon of life: Toward a philosophical biology*. Northwestern University Press.

Kim, E. J. (2021). Information geometry, fluctuations, non-equilibrium thermodynamics, and geodesics in complex systems. *Entropy*, *23*(11), 1393.

Koster, J. T. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics*, *26*(3), 413–431.

Lanillos, P., Meo, C., Pezzato, C., Meera, A. A., Baioumy, M., Ohata, W., … Tani, J. (2021). Active inference in robotics and artificial agents: Survey and challenges. *arXiv preprint*, arXiv:2112.01871.

Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv preprint*, arXiv:1801.00631.

Materassi, D., & Salapaka, M. V. (2014). Notions of separation in graphs of dynamical systems. *IFAC Proceedings Volumes*, *47*(3), 2341–2346. https://doi.org/10.3182/20140824-6-ZA-1003.02661

Mazzaglia, P., Verbelen, T., Çatal, O., & Dhoedt, B. (2022). The free energy principle for perception and action: A deep learning perspective. *Entropy*, *24*(2), 301.

Menary, R., & Gillett, A. J. (2020). Are Markov blankets real and does It matter? In D. Mendonca, M. Curado, & S. S. Gouveia (Eds.), *The philosophy and science of predictive processing* (pp. 39–58). Bloomsbury Academic.

Millidge, B., Seth, A., & Buckley, C. L. (2021). Predictive coding: A theoretical and experimental review. *arXiv preprint*, arXiv:2107.12979.

Millidge, B., Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). On the relationship between active inference and control as inference. In T. Verbelen, P. Lanillos, C. L. Buckley, & C. De Boom (Eds.), *Active Inference. IWAI 2020. Communications in Computer and Information Science* (Vol. *1326*, pp. 3–11). Springer. https://doi.org/10.1007/978-3-030-64919-7_1

Parr, T., Da Costa, L., Heins, C., Ramstead, M. J. D., & Friston, K. J. (2021). Memory and Markov blankets. *Entropy*, *23*(9), 1105.

Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological Cybernetics*, *113*(5), 495–513.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Pearl, J., Geiger, D., & Verma, T. (1989). Conditional independence and its representations. *Kybernetika*, *25*(7), 33–44.

Pearl, J., & Mackenzie, D. (2018). *The book of why*. Basic Books.

Ramstead, M., Friston, K., & Hipólito, I. (2020). Is the free-energy principle a formal theory of semantics? From variational density dynamics to neural and phenotypic representations. *Entropy*, *22*(8), 889.

Reed, E. S. (1989). Neural regulation of adaptive behavior: An essay review of neural Darwinism. *Ecological Psychology*, *1*(1), 97–117. https://doi.org/10.1207/s15326969eco0101_5

Reed, E. S. (1996). *Encountering the world: Toward an ecological psychology*. Oxford University Press.

Rosas, F. E., Mediano, P. A., Biehl, M., Chandaria, S., & Polani, D. (2020). Causal blankets: Theory and algorithmic framework. In *International workshop on active inference* (pp. 187–198). Springer.

Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cognitive Processing*, *17*(3), 279–305.

Wheeler, J. A. (1982). The computer and the universe. *International Journal of Theoretical Physics*, *21*(6–7), 557–572.

Wolfram, S. (2002). *A new kind of science*. Wolfram Media.

Zuse, K. (1982). The computing universe. *International Journal of Theoretical Physics*, *21*(6–7), 589–600.