A critical analysis of Markovian monism

Majid D. Beni¹

Received: 30 September 2020 / Accepted: 4 February 2021 / Published online: 16 February 2021
© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract

Free Energy Principle underlies a unifying framework that integrates theories of origins of life, cognition, and action. Recently, FEP has been developed into a Markovian monist perspective (Friston et al. in *BC 102*: 227–260, 2020). The paper expresses scepticism about the validity of arguments for Markovian monism. The critique is based on the assumption that Markovian models are scientific models, and while we may defend ontological theories about the nature of scientific models, we could not read off metaphysical theses about the nature of target systems (self-organising conscious systems, in the present context) from our theories of nature of scientific models (Markov blankets). The paper draws attention to different ways of understanding Markovian models, as material entities, fictional entities, and mathematical structures. I argue that none of these interpretations contributes to the defence of a metaphysical stance (either in terms of neutral monism or reductive physicalism). This is because scientific representation is a sophisticated process, and properties of Markovian models—such as the property of being neither physical nor mental—could not be projected onto their targets to determine the ontological properties of targets easily.

Keywords Free energy principle · Markovian monism · Fictional models · Mathematical models

1 Introduction

Following in footsteps of Hermann von Helmholtz, Friston and colleagues have launched a strong research program that subsumes theories of perception, cognition, action, and life under the umbrella of Free Energy Principle (Buckley et al.

I thank two anonymous reviewers of this journal for their insightful comments. Also, Wanja Wiese read an earlier draft of this paper and provided precious comments. All of these debts are gratefully acknowledged.

✉ Majid D. Beni
m Beni@metu.edu.tr

¹ Department of Philosophy, Middle East Technical University, Ankara, Turkey

2017; Constant et al. 2018; Friston, 2009; Friston et al. 2012). Free energy is an upper bound on surprise—which is the negative log probability of an outcome—given a generative model—which is a model of probabilistic dependencies between causes and consequences (Friston 2010, pp. 1–3). According to FEP, self-organising systems draw on approximate Bayesian inferences to reach non-equilibrium steady-state. Organisms resist the natural tendency to disorder by minimising the occurrence of surprising events, whose occurrence does not provide evidence for organisms’ generative models. Informally, this is stated by saying that self-organising systems are self-evidencing, in the sense that they garner evidence for their existence (Hohwy 2014). By staying in states with low surprise, organisms maximise their survival, and organisms mainly minimise their surprise by minimising variational free energy, which is an upper bound on surprise. In this fashion, FEP underpins a fundamental groundwork for explaining “how a biological system, exposed to random and unpredictable fluctuations in its external milieu, can restrict itself to occupying a limited number of states, and therefore survive in some recognisable form”(Friston 2012, p. 2100). FEP has been applied to the Bayesian theories of brain and cognition (Friston and Stephan 2007).

The philosophical implications of FEP have been discussed extensively by eminent philosophers (Clark 2016a; Gallagher and Allen 2016; Hohwy 2013). But FEP-theorists remained content with providing general promissory insights into the possible scientific basis of forthcoming philosophical solutions. The situation has changed somewhat dramatically in Friston et al. (2020) recent statement of a metaphysical perspective on consciousness, and as such their recent enterprise is worthy of admiration. Their offered metaphysical perspective is named Markovian monism. Markovian monism holds that Markov blankets are the type of thing/property that constitutes both mind and matter, and in this sense, their metaphysical monism is dubbed Markovian. In a brave statement of the metaphysical component of their thesis, Friston et al. (2020, p. 1) submit that “fundamentally, there is only one type of thing and only one type of irreducible property”. Markov blankets form the “calculus of belief” which is not only a nice scientific-mathematical model of consciousness but also provides a clear metaphysical statement and can shed light on “current philosophical debates” (ibid).

The virtues of the Markovian account of consciousness notwithstanding, it fails to support the monist stance with valid arguments and Friston et al. industrious use of information geometry (which will be discussed later in this paper) does not contribute to the construction of a solid ontological perspective on consciousness. The paper substantiates this critical point through an analysis of the relationship between scientific models—as the vehicle of scientific representations—on the one hand, and their targets—conscious self-organising systems, in this case—on the other hand. I construe ‘Markovian models’ as material/fictional/mathematical models and argue that none of these interpretations provides the requisite basis for defending a monist stance. The critique is stated in the hopes of motivating the FEP-theorists to improve the present statement of Markovian monism.

The paper is structured in the following way. In Sect. 2, I outline Markovian monism and show how an objective evaluation of Markovian monism requires pondering the question of the relationship between scientific models and their target systems.

In Sect. 3, I provide some preliminaries about scientific modelling. I also explain that scientific representation is usually an indirect relation, which does not warrant deriving conclusions about the ontological properties of target systems based on the properties of scientific models. Then I consider different ways of construing Markovian models. In Sect. 4, I construe Markovian models as material models but argue that this construal does not back up metaphysical monism. In Sect. 5 I construe Markovian models as fictional entities and reach the same conclusion as Sect. 4. In Sect. 6 I construe Markovian models as mathematical models and reach the same conclusion. We cannot read off ontological features of the target system (i.e., consciousness) from the ontological status of Markovian models. In Sect. 7, I assess Friston et al. remark on the affinity of Markovian monism with structural realism in the philosophy of science and argue that Markovian monism could not be defended in the spirit of structural realism. This demonstrates that Markovian monism cannot be defended as a form of (reductive) materialism. Section 8 demonstrates that Friston et al.'s possible defence of neutral monism is marred by a fallacy. Section 9 is the conclusion.

2 Modelling the origin of consciousness

FEP-theorists were toying with the philosophical question of what is consciousness (Hobson and Friston 2014; Solms and Friston 2018). And indeed an interesting theory of (self-)consciousness has evolved out of the theoretical framework of FEP. Almost all organisms can model the states of the world to themselves to some extent, but only some of them can be (self-)conscious. The difference between these two types of organisms (with and without consciousness) is that organisms that are capable of (self-)consciousness not only model their environment but can also model their *expectations* about the *future* states and *consequences* of their actions. Such organisms can form temporally *thick* or *deep* models of their environment (Friston 2018, p. 579). Such organisms can perform active inferences. The notion of 'active inference' models the distribution of posterior beliefs over action, under the prior belief that action will minimise free energy in future (Friston 2018, p. 279). This definition provides a basis for explicating intentionality and agency, in the sense that the relevant organisms can perform purposeful behaviour, and choose the right sort of policies for surviving in the world (Friston et al. 2010, 2015).

We are going to contextualise the role of Markov blankets in FEP shortly. But before that, I have to remark that there are various ways of construing the ontological implications of FEP. Some interpretations regard FEP as a fundamental hypothesis or theory with somewhat realist tendency to represent nomic relations between the brain and the environment (Allen and Friston 2016; Wiese and Metzinger 2017). However, it is also possible to argue that being an axiom or postulate that is grounded in theoretical physics and biology makes FEP a scientific principle (Colombo and Wright 2018). In this vein, it may be (and indeed has been) argued that FEP provides a principled explanation that serves as a guide to which things may or may not accord, rather than a hypothesis that provides a bona fide description

of target systems.¹ This alternative does not need to support realism, meaning that it does not assume that Bayesian networks or Markovian models that are invoked by FEP are representing features of the target systems. Below, I shall unpack ramifications of this point.

At least according to some interpretations, understanding FEP as a principle might pose a challenge to claims about FEP's explanatory power. For example, it could be contended that FEP basically amounts to tautologies and thus it is liable to the objection from triviality (Klein 2018). However, construing FEP as a mere principle that is devoid of realist implications does not necessarily need to undermine the explanatory power or scientific credentials of FEP. For example, it can still be assumed that the information-theoretic framework of FEP provides a common intellectual ground for researchers from various fields of expertise to offer piecemeal explanations of mechanisms of self-maintenance and self-organisation (Colombo and Wright 2018). As such, the scientific status of FEP will remain unscratched, even without tying the explanatory framework of FEP to realist interpretations. In fact, one does not even need to be a critic of FEP to appreciate the appeal of deflationary, non-realist interpretations of the FEP. For example, van Es and Hipolito (2020) confute both representationalist and non-representationalist varieties of a realist reading of FEP. Similarly, Ramstead et al (2020) have endeavoured to develop a non-realist construal of FEP and profess agnosticism about the ontic status of neural representations under FEP (more on this in next sections).

The short discussion in the previous paragraph does not intend to challenge the scientific value or explanatory prowess of FEP. It simply draws attention to the fact that there are different perspectives on ontological implications of FEP, and not all interpretations are in line with a realist characterisation of FEP. We shall attend to the role of Markov blankets in the statement of the FEP-based account of consciousness soon, but as we will shortly see (with reference to (Bruineberg et al. 2020)), the characterisation of Markov blankets too is prone to diverse ontological interpretations. Non-realist interpretations of Markov blankets, when substantiated adequately, provide grounds for scepticism about Markovian monism, which projects properties of the Markovian models to their target systems.

Be that as may, deep or thick models are at the centre of the FEP-based account of self-consciousness, which explains how some organisms can register the internal structure of the *subject* of the phenomenal experience (Friston 2018; Hobson and Friston 2014; Solms 2019; Solms and Friston 2018). Organisms with deep models can conceive of different consequences of their actions, and this bestows upon their self-evidencing processes a counterfactual richness. They could choose one particular course of action amongst several possible ones (Friston 2018, p. 6). In this vein, a self-conscious system is defined as “a system that can simulate multiple futures, under different actions, and select the action that has the least surprising outcome” (Friston 2018, p. 5).

There were insightful discussions of the philosophical implications of FEP (Bitbol and Gallagher 2018; Clark 2016b; Hohwy 2014). But the metaphysical question

¹ I thank one of the reviewers of this journal for bringing this point to my attention.

of the nature of consciousness have been rarely addressed in the literature. FEP-theorists modestly remained content with making general remarks on possible evolving solutions to the philosophical problems (Solms and Friston 2018), but they did not hold strong philosophical stances. An interesting straightforward FEP-based metaphysical account of consciousness has been presented only recently (Friston et al. 2020). It is dubbed Markovian monism, and it is worthy of a detailed examination.

The general insight behind Markovian monism is that dualism is orthogonal to the formal approach to modelling mental properties and consciousness by Markov blankets (Friston et al. 2020, p. 17). Friston et al. (2020) provide a detailed mathematical-Markovian formulation of the self-organising systems that can exhibit consciousness, and they conclude that “there is only one type of thing and only one type of irreducible property” that is modelled by Markov blankets (p.1). More importantly, according to Friston et al. formalism is not just an instrument of modelling what we think we may know about the target systems, but it tells something interesting about the *origins of mind and consciousness* (ibid, also see the title of their paper). What is lost in this glamorous elaboration of Markovian models of consciousness is that Markovian models are scientific models, and their relationship with their target systems is prone to philosophical interpretations.

The critical insight of the present paper is that we cannot infer the nature of a target system solely based on what we know about the nature of scientific models that represent it. More generally, the critical point can be stated in terms of a question; how could a scientific model (and a formal one at that) establish a metaphysical thesis (such as monism) or confute another thesis (such as dualism) about the ontological status of consciousness in self-organising systems. Interestingly enough, Friston et al. acknowledge that the issue of incompatibility of (property) dualism or lending support to monism is not only a matter of formalism but also a matter of “interpretation” (Friston et al. 2020, p. 18). This subtle remark opens up Pandora’s Box.

Markov blankets are configurations of spaces in Bayesian networks, where they connect nodes that form a blanket which separates a sub-space in the network from the rest of the network. A Markov blanket of a node consists of its parents, its children, and other parents of its children. States inside are independent of the states outside, given the relevant Markov blankets. Friston et al. (2020, Section 3) use Langevin formalism and Density Dynamics to weave Markov blankets into FEP and its account of prediction error minimisation. This formulation unifies basic insights of reinforcement learning, infomax, and the Bayesian brain theory (Friston et al. 2020, p. 7 ff.). Those fine details do not need to be reviewed here. Suffice it to say that Friston et al. construct their calculus of belief in terms of information geometry, and they use information length to measure the depth of temporal aspect of self-organisation and model the dynamical interaction between sensory states and active ones.

Friston et al. use information geometry to model the difference between ‘probability distribution *about* things’—which model belief states—and ‘probability distribution *of* things’—which model physical objects. In this fashion, their calculus could map belief states onto their target in the external world. The informational/probabilistic model of the relationship between ‘probability distribution *about* things’ and ‘probability distribution *of* things’ describes the way that the brain represents the external world to itself. It explains how the brain’s internal models are

coupled with their target systems in the external world. Markov blankets demarcate internal states from external states. Inside each Markov blanket, sensory states are separated from active states (where sensory states are independent of internal states whereas active states are independent of external states). However, it should be remarked that although internal and external states are independent, this independency is conditional to the extent that internal and external states indirectly influence each other in a loop via exchanges between active and sensory states² (where active states correspond to children of internal states and sensory states are represented as the parents of internal states). It is also worth mentioning that there is a disparity between various ontological interpretations of Markov blankets (this is similar and related to the aforementioned disparity between realist and non-realist interpretations of FEP). I draw on Bruineberg et al. (2020) recent study of Markov blankets to pinpoint this disparity.

Bruineberg et al. (2020) remarked the divergence between ontic interpretations of Markov blankets in terms of the difference between Pearl blankets and Friston blankets. Pearl blankets as being presented by Judea Pearl (1988), are powerful formal modelling tools that can regiment conditional independencies/separations in Bayesian networks (of considerable size). Friston blankets (Friston 2013) on the other hand are associated with an ontological perspective on the nature of the relationship between organisms and their environments. That is to say, Friston blankets aim to provide truth-apt (my term) descriptions of the patterns of dependencies and independencies between the organism and its environment. The ontological assumption here is that the properties of blankets can be projected onto their target systems. Despite being neatly supported by insights into cognitive psychology and evolutionary biology, the ontological insight that underlies Friston blankets would benefit from further criticism and discussion. Part of recent literature engages in this discussion by remarking that the metaphysically laden conception of Friston blankets cannot be justified only on the basis of the formalism of Markov blankets (Bruineberg et al. 2020, p. 19), (also van Es 2020; van Es and Hipolito 2020).

Let me recap quickly. According to Bruineberg et al.'s interpretation, there is a significant difference between Pearl's and Friston's respective characterisation of Markov blankets. Pearl's version does not look beyond the formalism itself. Friston's (2013) characterisation, on the other hand, projects some of the properties of the blankets onto their target systems. But the realist implications of this imputation/projection must be pondered cautiously because there is no monolithic ontological reading of the relationship between scientific models (such as Markovian models) and their target systems. The present paper develops a variation of this critical theme.

After regimenting the capacity of organisms to represent their environment in terms of information geometry, Friston et al. proceed to demonstrate that the same probabilistic/informational mapping could be invoked to model *expected surprise* as the set of beliefs that organisms hold about the consequences of their actions and their future. In this context, temporal aspect is represented by density

² I thank one of the reviewers of this journal for bringing this point to my attention.

dynamics of the intrinsic states of self-organising systems. By constructing an information-geometrical model of the temporal length of organisms' insights into the consequences of their actions, Friston et al. propose to provide a measure for consciousness; organisms with short temporal trajectories (or shallow temporal models) react reflexively to stimulus, whereas organisms with long (deep) temporal models could perform pre-mediated adaptive action. In this fashion, the Markovian approach could model intentionality, agency, consciousness and self-conscious states. The development of information geometry—consisting of adding a metric for measuring the distance to informational states space—and using it to measure free energy minimisation is a great feat. However, from the philosophical point of view, the most intriguing aspect of their 2020 paper is that it has been developed into a straightforward metaphysical stance on consciousness; namely Markovian monism. Friston et al. claim that finding the “right” structure and function is “metaphysically sufficient” for [identifying] consciousness (Friston et al. 2020, p. 20). But how could we ensure that we found the right structure, and why should identifying the structure be sufficient for identifying the metaphysical nature of consciousness?

My reservation about Markovian monism is this; Friston et al. submitted that for the mathematical equations to be adopted as models of the scientific phenomena, they must be interpreted in the right way (Friston et al. 2020, p. 18). A part of this process consists of pairing properties of the mathematical structure with the properties of the target system, by ascribing quantities and numerical values to the target system which embeds the phenomenon under study. Interpreting the mathematical structure and imputing it to the target system could take place quite formally, say, via partial isomorphism (more on this topic later). However, in order to develop this formal procedure into a defence of a metaphysical thesis, we must go well beyond the interpretation of formal structures and ponder the question of ontological properties of target systems (not ontological properties of scientific models). As an example, a small-world network (in terms of Watts and Strogatz 1998) could be invoked to predict the rate of transmission and scope of patterns of the spread of COVID-19 in the real world. We must interpret the vertices in the graphs as human beings and edges as patterns of social relations to pair properties of the small-world network with properties of human communities. However, the process of interpretation does not tell us much about the biochemical constitution of the viruses, or their ontological status, for that matter. The step from finding the adequate models of phenomena to making ontological claims is rather a long one.

3 On scientific models and representations

To a first approximation, it may be assumed that Markov Blankets are not modeler's tools but objective features of things with separable internal and external states. For example, Friston has recently declared that “The Markov blanket is not some statistical device by which we come to observe or model the world—it is a necessary attribute of a universe that can be carved into things” (Friston 2019, p. 176). But I think we can take Friston's words here with a pinch of salt. Markov blankets are not objective features of things or the universe but mathematical Bayesian models that

can be used to represent some (possibly) objective features of the universe. When elaborating on the mentioned view, Friston also clearly remarks that “in quantum mechanics, a Markov blanket is necessary to distinguish between the state of a system and the system performing a measurement. In statistical mechanics, the Markov blanket enables us to talk about thermodynamic systems in contact with a heat reservoir. Finally, in classical mechanics, Markov blankets are necessary to distinguish massive bodies that exert forces on each other” (ibid). But we know that if we scrutinise quantum states or states of a thermodynamic system or macroscopic states of bodies and forces, we do not find anything such as Markov Blankets there. Rather, we will find (almost) separable internal and external spaces that could be *modelled* efficiently by invoking Markov blankets. As such, Markov blankets, along with Bayesian models, Gaussian functions, and network models are modeller’s tools, and arguably, Friston et al. (2020) use Markov blankets to *model* features of self-organising systems. This means that Markov blankets are formal tools in the modeller’s toolbox. The modeller may use these tools to regiment the constellation of integration and optimisation of information. A similar point has been recently brought to light by Van Es and Hipolito (2020), who provide a detailed argument to the effect that the role of the modeller and her socio-cultural context must be taken into consideration when we speak of an organism’s ability to instantiate a model relative to its phenotype. The important question is, what kind of ontological conclusions we can derive from use of Markovian models in the context of the study of life and cognition? Before going further, I offer some platitudes about the nature of scientific modelling and scientific representations.

There are various philosophical stances on the relationship between models and their targets, in terms of direct or indirect representations. In this paper, I take it for granted that representations via scientific models are indirect.³ This view has been defended by Michael Weisberg (Weisberg 2007a, b) and Peter Godfrey-Smith (2006), among others. Weisberg (2013, 2007b) specifies modelling as the indirect way of studying and analysing the world. Frigg and Nguyen (2016) argue that indirect representation via scientific models consists of states of denotation, exemplification, keying up and imputation. For the sake of simplicity, I generally assume that the process of scientific representation includes stages of interpretation of models and their application to the target systems. Indirect representation is germane to modelling procedures. Modelling process includes a great amount of idealisation, approximation, and abstraction. Because scientific models represent the features of

³ There are also philosophers who assume that scientific models represent their targets directly, by enabling us to imagine the target of the model, in the same way that a statue or literary piece enable us to imagine a person (say, in statue of Napoleon) or a situation (Napoleonic wars, in *War and Peace*) respectively (Toon 2012). The important factor in spotting the relation between the model and its target is their similarity. A problem with this account of similarity is that, as I have argued elsewhere, in (Beni 2019 Section 2.5), it is too vague to provide a clear and detailed understanding of mechanisms of representation, at least in comparison with some other clearly stated accounts of scientific representation. Also, as Frigg and Naguyen (2016, p. 233), it is unclear how imagining that the target possesses certain properties on the basis of what the model says provides a foothold for attributing such features to the target system in an objective way. So, I assume that scientific representation is indirect.

reality indirectly (if at all), we cannot simply project properties of models to reality. I build upon this insight to show that the argument that presents Markovian monism as a metaphysically significant thesis is invalid. To the extent that I can see, Friston et al. (2020) may use Markovian models in one of these three ways;

- A. Markovian models are used in the sense of material entities.
- B. Markovian models are fictional entities.
- C. I think the most viable construal of Markovian models is to understand them as mathematical objects. This construal is more intriguing, not least because it lures us into the discussion of the relationship between mathematical models and the real systems that are represented by them. Friston et al. (2020) about structural realism, too, is in harmony with this line of interpretation. I shall speak extensively about this construal in this paper and relate it to the discussion of structural realism in Sect. 7 of this paper.

The general insight behind my critique is this: regardless of the ontological category which subsumes the models, models are surrogates that represent their targets (Swoyer 1991), and as such models cannot be the true subjects of ontological commitments. This point is elementary, but ignoring it could undermine otherwise interesting metaphysical projects. I shall unfold this point regarding different types of scientific models in the next three sections.

4 As physical models

Friston et al. (2020 Section 13) advocate a reductive materialist construal of consciousness. How could Markovian models be used to support reductive materialists? One possible (although not promising) way of doing this may consist of showing that Markovian models and their target systems are types of the same ontological category, namely physical objects. Below, I shall provide some examples of material models.

Phillips and Newlyn's hydraulic model of economy is a fine example of physical models (Morgan and Boumans 2004). A hydraulic pump is a physical object, and for the hydraulic model to work, properties of the economic systems must be identified with hydraulic properties of the pump. In this context, the interpretation of the model fixes the representational relation between a litre of water and the currency (in the monetary sense). Another good example is John Kendrew's plasticine model of myoglobin (Kendrew et al. 1958). Myoglobin is an oxygen-binding iron-binding protein in globular cells of many mammals. The concentration of myoglobin in red blood cells of some mammals such as whales make them capable of holding their breath for long periods. To represent the structure of myoglobin, Kendrew constructed a plasticine model. As its name indicate, the structure of myoglobin is moulded from a huddled rope of plasticine set upon few metal rods (de Chadarevian 2018, p. 1140). The epistemic/explanatory benefits of the model are quite remarkable. Although the modellers draw on the electron density data to construct the rope

model, the model itself is not simply a summary of these data. That is to say, the model is not simply “a tool to communicate effectively the information the data contained. The model provided epistemic access to the tertiary structure of the molecule in a way that the electron density data alone could not” (Frigg and Nguyen 2016, p. 226). However, despite its explanatory and predictive virtues, the rope model does not mean to indicate that myoglobin is actually made of plasticine. The question then is how serious are the ontological implications of the physical models? Let us apply this to our discussion of Markovian monism.

At times, FEP-theorists indeed imply that Markov blankets are physical models. For example, Ramstead et al. argue that Markovian models could be understood as “a physical transcription of causal regularities in its eco-niche that has been sculpted by reciprocal interactions between self-organisation and selection over time” (Ramstead et al. 2017, p. 5). Similarly, Pezzulo and Levin (2017) argue that Markov blankets are nested in the brain at the level of individual neurons, small neural networks and big functional networks. In the body, too, the Markov blankets are nested in the organisation of cells, organs and the whole body. For example, organic cells (in sensory epithelia) embody the function of the intermediate (sensory) states of the Markov blankets. (Pezzulo and Levin 2017, p. 2). More recently, (when discussing ideas from Allen and Friston (2016)) Bruineberg et al. (2020, p. 21) have argued that Markov blankets have been at times equated with physical boundaries that distinguish a system from its environment. These assertions may indicate that Markov blankets are physical entities. However, there are reasons for not assuming that Markovian models are physical objects. Firstly, as I will argue in the next two sections, to the extent that FEP-theorists’ remarks are concerned, Markovian models could be as well identified with fictional and mathematical entities (so, there is no consensus that they are physical objects). Moreover, almost everything with internal and external states could be modelled in terms of Markov blankets, and almost everything has internal and external states. It follows that all things (or at least all things that can be measured) are made of Markov blankets. I shall provide an example immediately.

Kirchhoff et al. (2018) provide a detailed Markovian model of a set of coupled Huygens’ pendulums which exemplify coupled random dynamical systems. The oscillating systems are suspended from a moving beam, which serves as a Markov blanket. In this setting, “motions of the two pendulums are statistically independent of one another conditioned on the motion of the beam. ... Each pendulum can be understood as a generative model of the other, where the probabilistic mapping from hidden causes (the dynamics of the black clock) to sensory observations (for the grey clock) is mediated by the beam, i.e. the Markov blanket states of the clocks” (Kirchhoff et al. 2018, p. 5). While Markov blankets do not seem to be physical models of Huygens’ pendulums, or they are nested in the clocks or are identical with oscillating systems suspended from moving rafters. However, according to Kirchhoff et al. using theories of optimisation to model the dynamical behaviour of the system in terms of a Markov blanket allows us to recognise the system as an agent that garners evidence for its existence. But then again, others such as Bruineberg et al. (2020 Sect. 6) contest the realist reading by pointing out that the reading conflates between a formal framework and a metaphysically laden interpretation of formalism

(they call these Pearl blankets and Friston blankets respectively, see Sect. 2 of this paper). Without intending to endorse Bruineberg et al.'s terminology, I just second their point that “realist reading cannot be justified by just following from the mathematics” (Bruineberg et al. 2020, p. 41). Instead, it would be more sensible to consider Markovian models as scientific models used by modellers to represent target systems that are not made of Markovian models (for discussion see van Es 2020).

At any rate, Markovian models are not physical objects, but even if Markovian models were material objects, we could not ensure that there is any ontological affinity between the nature of Markovian models, on the one hand, and their target systems, on the other hand. That is to say, we could not ensure that consciousness falls under the same ontological category that includes Markovian models, and thus we could not conclude that consciousness is physical too (obviously, such an argument, even when available, would have been vainly circular). In short, even if Markovian models had been material objects, we would not have been able to conclude that consciousness (as the target system) is material too. We still needed to demonstrate that consciousness is made of the same material stuff that (in a possible scenario) constituted Markov blankets. However, not only Markov blankets are not material things, there is no reason to assume that consciousness would share the ontic nature of Markov blankets, even if Markovian models were physical.

5 As fictional models

Let us assume that Markov Blankets are not concrete physical objects. It could be incontrovertibly confirmed that if we explore actual brains of organisms we do not find any actual, concrete Markov Blankets ingrained in their flesh and blood. But we could still assume that Markov blankets are fictional objects. There are various ways of conceiving of scientific models as fictional objects (Frigg 2010; Frigg and Nguyen 2016; Salis 2019).

According to one approach, developed in (Toon 2012) amongst other places, scientific models are just props in the game of make-believe. These props can be used to mandate those who engage in the game to imagine certain things about the target system, without intending to convey any metaphysical implications about the target systems. Under the fictionalist approach, “[r]epresentation is explained in terms of there being the prescription to imagine certain things about the target [...]. At the same time vehicles, understood as ‘secondary systems’ [i.e., models], are rendered otiose, which dissolves any metaphysical questions about these systems” (Frigg and Nguyen 2016, p. 233). Thus, ontological features of the target systems cannot be inferred on the basis of scientific models. Another statement of fictionalism, developed in Contessa’s (2010) work, indicates that like fictions, scientific models have an author (or modeller), meaning that they are constructed by a modeller for specific purposes. Moreover, at least some scientific models have characteristics that are similar to characteristics of physical models, but unlike material models, fictional models do not exist. Sherlock Holmes was said to be a smoker, without actually existing (Contessa 2010, p. 218). In the scientific context too, models may serve as non-existing things that describe properties of things that exist. On such grounds,

Contessa argues that fictional characters from novels and fictional scientific models are both types of the same ontological kind, namely imaginary possible objects (Contessa 2010, p. 219). In this light, scientific assertions, say about the existence of electrons, are to be understood as parts of a fiction that serves to produce useful empirical results. Let us see how fictionalism applies to Markovian monism.

There is some textual evidence for construing Markovian models of consciousness as fictional models. For example, when using Langevin formalism to characterise Markovian models, Friston et al. submit that if we model adaptive behaviour of a creature by Markovian blankets, “it will *look as if* it is acting to minimise its particular surprisal (or variational free energy). In other words, it will *look as if* it is trying to minimise the surprisal, expected following an action” (Friston et al. 2020, p. 13 emphasis added). Markov blankets could be deployed as ingredients of fictions to produce empirically adequate models of self-consciousness. On the same note, Ramstead et al. (2020) have recently defended a non-realist interpretation of the explanatory role of neural representations under FEP. This interpretation considers neural representations as fictions that provide a scientifically useful way of describing mechanisms of cognition and action, without making any commitments to the reality or existence of such mechanisms. In the same vein, van Es and Hipolito (2020) submit a detailed examination of both representational and non-representational interpretations of the functioning of generative models to argue that neither of these interpretations supports a realist reading of FEP. (Interestingly enough, this is because the realist interpretation ignores the role of agents—or modellers—who manipulate Bayesian models and project them to target systems (van Es and Hipolito 2020 Section 3.1)). On a more positive note, elements of FEP could be construed along the lines of an outright instrumentalist approach which recognises the explanatory power of FEP without burdening it with problematic realist assumptions. The general insight behind the endorsement of instrumentalism is that “from the fact that it is possible to model a process, it does not necessarily follow that the target phenomenon represents the intellectual tools we use to model it” (van Es and Hipolito 2020, p. 21). Below, I shall apply this insight to Markovian monism.

Let us assume that Markovian models, as elements of FEP, are indeed pieces of useful scientific stories that make the operation of cognitive systems understandable. But, and instrumentalist-fictionalist reading of Markov blankets hardly supports Markovian monism. That is to say, the consequences of considering Markovian models of consciousness as fictional models would be rather dire for Markovian monism. If Markovian models are fictional entities, we could not use them to infer the ontological nature of consciousness, which, allegedly, is not a fictional entity (in the same way that we could not infer much about the ontic status of any actual divine being based on the description of Greek gods in Hesiod’s *Theogony* (assuming that it is a work of fiction)). The fictionalist construal of Markovian models does not contribute to cementing monism about consciousness. If Markovian models are just useful fictions or efficient modelling tools in the modeller’s toolbox, then applying them to provide empirically adequate representations of target systems does not warrant deriving conclusions about the ontological status of the target systems. Thus Markovian monism cannot receive the requisite support, and it would remain a groundless metaphysical thesis.

6 As mathematical models

Markovian models do not seem to be material objects, and construing them as fictional entities does not support monism about consciousness either. Finally, we may consider Markovian models as mathematical objects. There is ample textual evidence for this construal. Friston et al. (2020, p. 1) submit that their enterprise aims to:

describe constructs from information theory and physics that place certain constraints on the dynamics of self-organising creatures, such as ourselves. These constraints lend themselves to an easy interpretation in terms of beliefs and intentions; provided one defines their meaning carefully about the mathematical objects at hand.

Mathematical models are ubiquitous in sciences, as modern science widely draws on mathematical formalism to describe natural phenomena. In the philosophy of science, too, there is a strong tendency to invoke set/model-theoretic constructs to represent natural phenomena (Suppes 1967; van Fraassen 1970).

A conspicuous example of mathematical models is Volterra's model of fluctuation of predator–prey dynamics (Volterra 1926). The predator–prey dynamics consists in the coexistence between two associated biological species, where the individual members of one of the species feed upon the members of the other species. Fishing had ceased in the Adriatic Sea during WWI. The number of fishes was supposed to increase as a result of the suspension of fishing during the war. In reality, though, the opposite was the case. Volterra's model can explain the phenomenon. It holds that “the proportional rate of increase of the eaten species diminishes as the number of individuals of the eating species increases, while the augmentation of the eating species increases with the increase of the number of individuals of the eaten species” (Volterra 1926, p. 558). I follow Weisberg (2007b, p. 210) to state the formula in terms of the dynamics of the relationship between the size of the prey population (V) and abundance of predators (P). In the equation, r stands for the growth rate preys and m is the death rate of the predators' population.

$$\frac{dV}{dt} = rV - (aV)P$$

$$\frac{dP}{dt} = b(aV)P - mP$$

This mathematical model explained the decrease of fishes and predicted that resumption of heavy fishing would lead to an increase in the number of fishes in the mentioned context (the post-war Adriatic Sea). The model accommodates plausible explanations (of the decrease in the prey population) and provides novel predictions (that heavy fishing will lead to augmentation of fishes). In light of the model's explanatory power and its innovative prediction, it could be granted that this model describes the ‘*right*’ sort of mathematical structures. (As I remarked before, Friston et al. assume that finding the right structure and function is enough for the

metaphysical identification of consciousness). How does that bear on the question of the ontology of sea creatures?

Differential equations are mathematical objects. Relationship between models and their targets could be regimented in fine mathematical terms too, namely in terms of (partial) isomorphism (da Costa and French 2003; van Fraassen 1980). There are mathematical realists (such as Pythagoreans and Platonists) who embrace metaphysical realism about mathematical entities and structures, and there are also philosophers who deny that mathematical objects exist (Field 1989). However, although not irrelevant, realism/antirealism about mathematical objects is not directly related to scientific realism. That is to say, in the context of Volterra's model, it is the existence of fishes and sharks (as the targets of the models), rather than the existence and properties of mathematical objects, that matter to the scientific realist. From the scientific realist point of view, ontological commitments should not be made to mathematical entities such as differential equations but to sea creatures and perhaps to natural kinds and categories that subsume sea creatures. If realism could be professed in the present context at all, it would endorse ontological commitments to entities such as fishes and sharks (in the first place) and perhaps commitments to natural kinds of preys and predators. In this context, scientific realism (or antirealism, for that matter) is concerned with the ontic status of the individual organisms (and perhaps natural kinds of preys and predators) which could not be derived from what our mathematical models of phenomena divulge about their target systems.

To return Markovian monism, Markov blankets may indeed provide the 'right' sort of models for consciousness in self-organising systems. But Markovian models are mathematical objects—it is indeed why Markovian models are neither mental nor physical models—and they do not provide any insight into the metaphysical nature of consciousness. Interestingly enough, Friston et al. seem to be anticipating this point. After outlining their information geometry, which grounds Markovian Models, they (2020, p. 17) ask does the formalism indicate that “all systems with a Markov blanket have a mind [...]? Are such systems conscious?” and to this, they answer that “The formalism itself does not answer these questions: different metaphysical interpretations of the existence of a dual information sssgeometry are possible. In fact, one might ask whether it has any metaphysical significance whatsoever” (ibid). This scepticism, if genuine, would run against the grain of Friston et al. (2020, p. 17) optimism about the implications of their theory about the origins of mind and consciousness, or its success in dealing with the hard problem.

Before going further, let me provide another textual evidence for a deflationary understanding of FEP which identifies variational Bayesian networks as mathematical models. Ramstead et al. (2020) (whose fictionalist approach has been alluded to in Sect. 5 of this paper) develop a variation of their instrumentalism into a deflationary account that specifies neural representations in terms of mathematical models. Although this deflationary account does not deny the existence of neural representations, it takes an overly antirealist attitude towards their content, where it is assumed that representations are not individuated by their content but by the mathematical functions that realize them (Ramstead et al. 2020, p. 12). According to Ramstead et al. there is a dark side to this deflationary reading, which (obviously) trivialises cognitive content and reduces its explanatory role to a heuristic one (this is because

the content can help the scientists to make sense of mathematical descriptions of various mechanisms and normal conditions that rule over them (Ramstead et al. 2020, p. 14)). In my opinion (and possibly in Ramstead et al.) this problem of the deflationary reading is not grave enough to discredit it. But again, if this construal of the representational mechanisms of FEP and its ingredients is correct, the consequences for Markovian monism could be rather grave. Given the general non-realist tendency of the deflationary account, the approach indicates that ontic status of minds and their phenomenal features could not be just directly read off from the mathematical formalism that constitutes the essence of FEP. That is to say, even if we grant that mathematical functions are reliable enough for fleshing out semantic aspects of cognition, we cannot infer much from them about the nature of the consciousness. Let me elaborate.

When interpreted appropriately, Markovian models ascribe some structure or function to consciousness. As Friston et al. (2020, p. 20) acknowledge, this may indicate that the FEP-based account fails to deal with the hard problem of consciousness, which demands something more than the specification of the structure of consciousness. But after acknowledging this, they immediately add that “if consciousness is a vague concept (as suggested by our interpretation of Markovian monism), then the right structure and functions can be metaphysically sufficient for consciousness...” (ibid). But even if we could find the right structure of the class of mathematical models that represent consciousness, we would be none the wiser the metaphysical status of consciousness. I will elaborate on this point in the next section, with an eye to Friston et al.’s reference to structural realism.

7 Structural realism

In the conclusion of their paper, Friston et al. remark that their Markovian monism is compatible with Structural Realism (SR). SR is a successful theory of contemporary philosophy of science which indicates that epistemological or/and ontological commitments are to be made based on the structure of theories, not their form (Ladyman 1998; Worrall 1989). There is not enough space in this paper to review philosophical motives that led to the formation of SR. Suffice it to say that structural realists usually build upon set/model-theoretic structures to regiment their conception of theories and scientific representations (Bueno et al. 2002; French 2006). That is to say, structural realists assume that model theory provides the *right* formal framework for regimenting the structure of scientific theories. Friston et al.’s development of information geometry and statement of the criterion of information length make it especially compatible with SR.⁴ However, finding the *right* formal framework is not enough for dealing with the question of the metaphysical status of the target systems.

⁴ Informational spaces do not by default include a geometry or the notion of distance. To incorporate geometry into information theory, Friston et al. (2020, p. 17) have supplied their information theory with “a metric tensor—such that small displacements in state space can be associated with a metric of distance.” The result is a Riemannian geometry of statistical manifolds which embeds the Gaussian probability density of the FEP-based models of belief states and causes of beliefs in the external states. This approach is quite in line with SR and its strategy of structural regimentation of representations.

There should exist an independent argument for substantiating the metaphysical claim. Friston et al.'s main argument for establishing the metaphysical component of their thesis rests on the distinction between the intrinsic and the extrinsic types of information geometry. The former kind of information geometry is “inherent in the information length based upon time-dependent probability densities over internal states. This information length characterises the system or creature in terms of itinerant, self-organising density dynamics that forms the basis of statistical mechanics in physics, i.e., a physical, material, or *mechanical* information geometry that is *intrinsic* to the system” (Friston et al. 2020, p. 11). On the other hand, extrinsic information geometry refers to belief distributions over external states. Part of Friston et al.'s grounds for Markovian monism rests on the assumption that the intrinsic information geometry is prior to the extrinsic information geometry, in the following sense: in order to interpret internal states as representations of probability distributions (over external states, given blanket states), one actually has to consider properties of ensembles of internal states (e.g., the average activity of a neural population). This presupposes that a more fine-grained description of internal states—which only describes probabilities of internal states, but not probabilities represented by internal states—is available, given that “It is tenable to associate physics (in the sense of quantum, statistical and classical) mechanics with the intrinsic information geometry. Indeed, this is common parlance in statistical physics” (Friston et al. 2020, p. 25). Accordingly, one could argue that the more fine-grained description picks out things that are ontologically more fundamental (Wiese, personal communication). I can see the charm of using the intrinsic information geometry to substantiate Markovian monism (given that this geometry may provide a fine-grained description of fundamental ontological units). However, the claim is open to debate in the following sense.

In the case of SR, structural realists extensively use a model-theoretic conception of scientific theories on grounds that the model-theoretic approach “wears its structuralist commitments on its sleeves” (French 2014, p. vii). However, this does not mean that metaphysical claims of structural realism are dependent on the choice of model-theoretic framework. Arguably it is possible to use other formal frameworks such as category theory (Halvorson 2012) or first-order logic (French 2014 Section 2.9 on Ramsey sentence) to regiment scientific representations. The choice of the model-theoretic structures as the right sort of vehicle for regimenting the scientific representations does not substantiate the ontological thesis of SR, which is supposed to be substantiated on independent grounds (in the philosophy of science, there are different ways of fleshing out such independent grounds. For example, it could be assumed that if the representations were dislodged from reality, the remarkable empirical success of scientific theories would have been a matter of coincidence in cosmic scales (Smart 1963). Because this is not the case, scientific theories are truth-apt and veridical). Although structural realists use set/model-theoretic structures to regiment scientific representations, they are not epistemologically/ontologically committed to mathematical structures. If they were committed to mathematical structuralism, they had to advocate some sort of Pythagorean or Platonist ontology, which is not the case (French 2014, p. 10). Rather they mark a distinction between the representational role of mathematical structures and the

ontologically constitutive role of physical structures, and they make their ontological commitments to physical structures, not formal structures (French 2006; Ladyman 2007). Therefore, although sometimes it has been suggested that structural realists are concerned with the ontology and epistemology of scientific models (Contessa 2010, p. 216), SR only uses set/model-theoretic structures for representational purposes. Ontological commitments to physical structures are supposed to be backed up by independent reasons (see French 2011). Back to the context of FEP, Friston et al. (2020) reliance on information geometry makes their theory particularly compatible with a structural realist reading. However, the choice of the intrinsic information geometry does not provide any insight into the mental/physical/neutral nature of consciousness. While we may procure independent philosophical arguments for any of these stances, Friston et al.'s fine development of Markovian models and information geometry does not contribute to establishing metaphysical monism.

8 A cartesian fallacy?

Despite what Friston et al.'s claim about the compatibility of their view with reductive materialism, I think it could be matched best with neutral monism. In fact, their paper also includes some indication of the incompatibility of Markovian monism with physicalism. For example, in the conclusion of their paper, they argue that:

Note, the claim here is that physics (i.e., statistical thermodynamics) supervenes on the same statistical manifold as belief updating. This supervenience—on the same statistical manifold—from which both information geometries inherit their structure could be read as the philosophical formulation of the mathematical conjugacy implied by intrinsic and extrinsic information geometries. (Friston et al. 2020, p. 24).

This means that Markovian models do not supervene on physical structures and are not reducible to them. Rather, physical structures are reducible to the neutral mathematical structures that underpin both physical and mental properties but are neither material nor mental. Therefore, Markovian monism could hardly be reconciled with any standard form of reductive materialism, and it meshes more nicely with neutral monism. But unfortunately, there is no good argument for associating Markovian monism with neutral monism either. Friston et al. do not concede a neutral monist reading of their theory, because they think that neutral monism mandates commitments to realism and panpsychism. I do not think these are good reasons for defying neutral monism. The source of my scepticism about a neutral monist reading of Friston et al.'s theory is something else. Below, I will draw attention to a fallacy that is inherent to Friston et al.'s defence of monism.

A possible defence of a neutral monist stance could be outlined in the following manner:

There are two conjugate ways in which things that exist can be described: either from the perspective of the intrinsic information geometry or from the perspective of the extrinsic information geometry. Under the assumption

that neither perspective is privileged, one would have to conclude that reality is, fundamentally, ontologically neutral. (Friston et al. 2020, p. 18).

But the argument is invalid. This is because from the fact that our formal descriptions are silent about the (mental/material) nature of specific things, it does not follow that those things do not fall under one of the categories of the mental or the material (or both!). Ironically enough, Descartes partly built his substance dualism on a similar invalid argument. In *A Discourse on the Method of Correctly Conducting One's Reason and Seeking Truth in the Sciences*, Descartes ([1637] 2008, p. 29) concludes that his essence (or the essence of any human being, for that matter) is thinking and a thinking being is not dependent on material things. Descartes argument is that while he knows he is thinking (or doubting), he is unaware of any connection between his thoughts and bodily mechanisms that may underpin them. But the problem is that from the ignorance of the underpinning physical mechanisms of thought, it does not follow that such mechanisms do not exist or thinking is independent of such mechanisms. Similarly, in the context of Markovian monism, from the fact that our formal description of energy minimisation is neither formal nor material, it does not follow that consciousness and mental properties are neutral commodities.

9 Concluding remark

The Free Energy Principle not only underpins interesting accounts of life, cognition, and action but arguably it could also lead to a fascinating theory of (self-) consciousness of self-organising systems. Recently, Friston et al (2020) have endeavoured to develop this insight into Markovian monism, which presumably provides a metaphysical stance on consciousness. In this paper, I expressed scepticism about the validity of Markovian monism as a plausible metaphysical thesis. To substantiate my point, I discussed different ways in which Markov Blankets could represent or model consciousness. I construed Markovian models as material models, fictional models, and mathematical ones, and argued that none of these interpretations lends support to a metaphysical perspective on consciousness in terms of either neutral monism or reductive materialism. As such, despite its scientific merits, the present statement of Markovian monism hardly leads to a viable metaphysical theory of consciousness or a good reply to the hard problem.

References

- Allen, M., & Friston, K. J. (2016). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese*. <https://doi.org/10.1007/s11229-016-1288-5>.
- Beni, M. D. (2019). *Cognitive structural realism: A radical solution to the problem of scientific representation*. Berlin: Springer.

- Bitbol, M., & Gallagher, S. (2018). The free energy principle and autopoiesis: Comment on “Answering Schrödinger’s question: A free-energy formulation” by Maxwell James Désormeau Ramstead, Paul Benjamin Badcock, Karl John Friston. *Physics of Life Reviews*.
- Bruineberg, J., Dolega, K., Dewhurst, J., & Baltieri, M. (2020). *The Emperor’s New Markov Blankets*. 1–47.
- Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, 81, 55–79. <https://doi.org/10.1016/J.JMP.2017.09.004>.
- Bueno, O., French, S., & Ladyman, J. (2002). On representing the relationship between the mathematical and the empirical. *Philosophy of Science*, 69, 497–518.
- Clark, A. (2016a). *Surfing uncertainty*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190217013.001.0001>.
- Clark, A. (2016b). Busting out: Predictive brains, embodied minds, and the puzzle of the evidentiary Veil. *Noûs*. <https://doi.org/10.1111/nous.12140>.
- Colombo, M., & Wright, C. (2018). First principles in the life sciences: the free-energy principle, organicism, and mechanism. *Synthese*. <https://doi.org/10.1007/s11229-018-01932-w>.
- Constant, A., Ramstead, M. J. D., Veissière, S. P. L., Campbell, J. O., & Friston, K. J. (2018). A Variational approach to niche construction. *Journal of the Royal Society, Interface*, 15(141), 20170685. <https://doi.org/10.1098/rsif.2017.0685>.
- Contessa, G. (2010). Scientific models and fictional objects. *Synthese*, 172(2), 215–229. <https://doi.org/10.1007/s11229-009-9503-2>.
- da Costa, N. C. A., & French, S. (2003). *Science and Partial Truth*. Oxford University Press. <https://doi.org/10.1093/019515651X.001.0001>.
- de Chadarevian, S. (2018). John Kendrew and myoglobin: Protein structure determination in the 1950s. *Protein Science*, 27(6), 1136–1143. <https://doi.org/10.1002/pro.3417>.
- Descartes, R. (2008). *A discourse on the method of correctly conducting one’s reason and seeking truth in the sciences* (I. Maclean (ed.)). Oxford University Press. <https://global.oup.com/academic/product/a-discourse-on-the-method-9780199540075?cc=ir&lang=en&>.
- Field, H. (1989). *Realism, mathematics, and modality*. Blackwell.
- French, S. (2006). Structure as a weapon of the realist. *Proceedings of the Aristotelian Society (Hardback)*, 106(1), 170–187. <https://doi.org/10.1111/j.1467-9264.2006.00143.x>.
- French, S. (2011). Metaphysical Underdetermination: why worry? *Synthese*, 180(2), 205–221. <https://doi.org/10.1007/s11229-009-9598-5>.
- French, S. (2014). *The Structure of the World: metaphysics and representation*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199684847.001.0001>.
- Frigg, R. (2010). Models and fiction. *Synthese*, 172(2), 251–268. <https://doi.org/10.1007/s11229-009-9505-0>.
- Frigg, R., & Nguyen, J. (2016). The fiction view of models reloaded. *The Monist*, 99(3), 225–242. <https://doi.org/10.1093/monist/onw002>.
- Friston, K. J. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, 13(7), 293–301. <https://doi.org/10.1016/j.tics.2009.04.005>.
- Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>.
- Friston, K. J. (2012). A free energy principle for biological systems. *Entropy*, 14(11), 2100–2121. <https://doi.org/10.3390/e14112100>.
- Friston, K. J. (2013). Life as we know it. *Journal of The Royal Society Interface*, 10(86), 20130475.
- Friston, K. J. (2018). Am I self-conscious? *Frontiers in Psychology*, 9, 579. <https://doi.org/10.3389/FPSYG.2018.00579>.
- Friston, K. J. (2019). Beyond the desert landscape. In Andy Clark & His Critics (Eds.), *Matteo Colombo, Elizabeth Irvine, & Mog Stapleton* (pp. 174–190). Oxford: Oxford University Press.
- Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological Cybernetics*, 102(3), 227–260. <https://doi.org/10.1007/s00422-010-0364-z>.
- 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. <https://doi.org/10.1080/17588928.2015.1020053>.
- Friston, K. J., Samothrakis, S., & Montague, R. (2012). Active inference and agency: Optimal control without cost functions. *Biological Cybernetics*, 106(8–9), 523–541. <https://doi.org/10.1007/s00422-012-0512-8>.

- Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159(3), 417–458. <https://doi.org/10.1007/s11229-007-9237-y>.
- Friston, K. J., Wiese, W., & Hobson, J. A. (2020). Sentience and the origins of consciousness: From cartesian duality to Markovian monism. *Entropy*, 2020(22), 516–516. <https://doi.org/10.3390/E22050516>.
- Gallagher, S., & Allen, M. (2016). Active inference, enactivism and the hermeneutics of social cognition. *Synthese*. <https://doi.org/10.1007/s11229-016-1269-8>.
- Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology & Philosophy*, 21(5), 725–740. <https://doi.org/10.1007/s10539-006-9054-6>.
- Halvorson, H. (2012). What scientific theories could not Be*. *Philosophy of Science*, 79(2), 183–206. <https://doi.org/10.1086/664745>.
- Hobson, A. J., & Friston, K. J. (2014). Consciousness, Dreams, and Inference: The Cartesian theatre revisited. *Journal of Consciousness Studies*, 21(1–2), 6–32.
- Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199682737.001.0001>
- Hohwy, J. (2014). The self-evidencing brain. *Noûs*, 50(2), 259–285. <https://doi.org/10.1111/nous.12062>.
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610), 662–666. <https://doi.org/10.1038/181662a0>.
- Kirchhoff, M. D., Parr, T., Palacios, E., Friston, K. J., & Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *Journal of the Royal Society, Interface*, 15(138), 20170792. <https://doi.org/10.1098/rsif.2017.0792>.
- Klein, C. (2018). What do predictive coders want? *Synthese*, 195(6), 2541–2557. <https://doi.org/10.1007/s11229-016-1250-6>.
- Ladyman, J. (1998). What is structural realism? *Studies in History and Philosophy of Science Part A*, 29(3), 409–424. [https://doi.org/10.1016/S0039-3681\(98\)80129-5](https://doi.org/10.1016/S0039-3681(98)80129-5).
- Ladyman, J. (2007). On the identity and diversity of objects in a structure. In *Proceedings of the Aristotelian Society, Supplementary Volumes* (Vol. 81, pp. 23–43). <https://doi.org/10.2307/20619100>
- Morgan, M. S., & Boumans, M. J. (2004). *Secrets Hidden by Two-Dimensionality: The Economy as a Hydraulic Machine*.
- Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. In *Probabilistic reasoning in intelligent systems*. Elsevier. <https://doi.org/10.1016/c2009-0-27609-4>
- Pezzulo, G., & Levin, M. (2017). Embodying Markov blankets: Comment on “Answering Schrödinger’s question: A free-energy formulation” by Maxwell James Désormeau Ramstead et al. *Physics of Life Reviews*. <https://doi.org/10.1016/j.plrev.2017.11.020>.
- Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2017). Answering Schrödinger’s question: A free-energy formulation. *Physics of Life Reviews*. <https://doi.org/10.1016/j.plrev.2017.09.001>.
- Ramstead, M. J. D., Friston, K. J., & 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. <https://doi.org/10.3390/e22080889>.
- Salis, F. (2019). The new fiction view of models. *The British Journal for the Philosophy of Science*. <https://doi.org/10.1093/bjps/axz015>.
- Smart, J. J. C. (1963). Philosophy and Scientific Realism. In *ROUTLEDGE & KEGAN PAUL*.
- Solms, M. (2019). The hard problem of consciousness and the free energy principle. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2018.02714>.
- Solms, M., & Friston, K. J. (2018). How and why consciousness arises: Some considerations from physics and physiology. *Journal of Consciousness Studies*, 25(5–6), 202–238.
- Suppes, P. (1967). What is a scientific theory? In S. Morgenbesser (Ed.), *Philosophy of Science Today* (pp. 55–67). Basic Books. https://www.google.com/_chrome/newtab?espv=2&ie=UTF-8.
- Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese*, 87(3), 449–508. <https://doi.org/10.1007/BF00499820>.
- Toon, A. (2012). Similarity and scientific representation. *International Studies in the Philosophy of Science*, 26(3), 241–257. <https://doi.org/10.1080/02698595.2012.731730>.
- van Es, T. (2020). Living models or life modelled? On the use of models in the free energy principle. *Adaptive Behavior*. <https://doi.org/10.1177/1059712320918678>.
- van Es, T., & Hipolito, I. (2020). *Free-Energy Principle, Computationalism and Realism: a Tragedy*.
- van Fraassen, B. C. (1970). On the extension of beth’s semantics of physical theories. *Philosophy of Science*, 37(3), 325–339. <https://doi.org/10.1086/288311>.

- van Fraassen, B. C. (1980). *The Scientific Image*. Oxford University Press. <https://doi.org/10.1093/0198244274.001.0001>
- Volterra, V. (1926). Fluctuations in the abundance of a species considered mathematically. *Nature*, 118(2972), 558–560. <https://doi.org/10.1038/118558a0>.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684), 440–442. <https://doi.org/10.1038/30918>.
- Weisberg, M. (2007a). Forty years of ‘the strategy’: levins on model building and idealization. *Biology & Philosophy*, 21(5), 623–645. <https://doi.org/10.1007/s10539-006-9051-9>.
- Weisberg, M. (2007b). Who is a Modeler? *The British Journal for the Philosophy of Science*, 58(2), 207–233. <https://doi.org/10.1093/bjps/axm011>.
- Weisberg, M. (2013). *Simulation and similarity using models to understand the world*. Oxford University Press.
- Wiese, W., & Metzinger, T. (2017). *Vanilla PP for Philosophers: A Primer on Predictive Processing*. <https://doi.org/10.15502/9783958573024>
- Worrall, J. (1989). Structural realism: the best of both worlds? *Dialectica*, 43(1–2), 99–124. <https://doi.org/10.1111/j.1746-8361.1989.tb00933.x>.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.