Journal for General Philosophy of Science
https://doi.org/10.1007/s10838-024-09673-w

ARTICLE

Structural Realism About the Free Energy Principle, the Best of Both Worlds

Majid D. Beni[1]

Accepted: 3 January 2024

Abstract
There are realist and antirealist interpretations of the free energy principle (FEP). This paper aims to chart out a structural realist interpretation of FEP. To do so, it draws on Worrall's (Dialectica 43(1–2): 99–124, 1989) proposal. The general insight of Worrall's paper is that there is progress at the level of the structure of theories rather than their content. To enact Worrall's strategy in the context of FEP, this paper will focus on characterising the formal continuity between fundamental equations of thermodynamics—such as Boltzmann's equation and Gibbs's equation—on the one hand, and Friston's characterisation of FEP on the other. Lack of a universal consensus on the physical character of entities that feature in thermodynamics, information theory and FEP notwithstanding, I argue that there is structural continuity and unity at the level of mathematical equations that regiment entropy, information and free energy. The existence of such structural continuity and unity provides grounds for structural realism about FEP.

Keywords Structural realism · Free energy principle · Scientific progress · Real patterns · Entropy

1 Introduction

The Free Energy Principle (FEP) has been developed by Karl Friston and colleagues (Friston 1994; 2010; Palm et al. 2015). There have been interesting philosophical debates about FEP's vast unifying scope and explanatory power (Clark 2016; 2020; Hohwy 2013). More recently, philosophical debates have been directed at realist or antirealist interpretations of FEP. In general, scientific realists make semantic, epistemic, and ontological commitments to the scientific description of worldly entities and structures. In this context, 'entity' refers to individual objects that endure over time. In contrast, structures are characterized by their relational nature. Now, realists about FEP accept semantic, epistemic, and ontological commitments to the free energy models that represent the mechanisms of perception, cognition, adaptive behaviour and so on. Antirealists, on the other hand, discard realist commitments. Instead of trying to arbitrate between antirealist and realist interpretations of FEP,

✉ Majid D. Beni
mbeni@metu.edu.tr

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

this paper aspires to find a stance that has the best points of realism and antirealism on its side; it will construe FEP along the lines of structural realism. Remarks have been already made on a structural realist reading of FEP. Friston et al. (2020), for example, remark that their formalism of FEP—in terms of information geometry—is in line with structural realism. However, comprehensive grounds for defending a structural realist interpretation of FEP have not been fully explored. This paper aims to address this gap. The paper's strategy for advocating a structural realist perspective of FEP is inspired by John Worrall's (1989) original articulation of structural realism.

In recent decades, Structural Realism (SR) has made significant contributions to various critical discussions, including those concerning the nature of individual entities (Ladyman 2007) and the challenge of metaphysical underdetermination (French 2006; 2011a; Ladyman 1998). This paper, however, is primarily concerned with a version of SR that aims to defend the existence of scientific progress, even in the presence of radical theoretical changes. It does so by emphasizing the continuity at the level of the underlying structure of theories. This version of SR has been advocated by Worrall (1989), and it holds that a viable position on scientific representation must have the best points of both realist and antirealist arguments on its side.[1] Worrall's proposal accounts for the progress of science at the level of the structure of theories across radical theoretical changes. This strategy presents SR in terms of the best of the realist and antirealist worlds.

Worrall's best-of-both-worlds strategy has been proposed in response to a long-standing division of opinion between scientific realists and antirealists, particularly within the field of philosophy of physics. A similar schism exists in the philosophical interpretations of the Free Energy Principle (FEP). On one side, there are realist interpretations that embrace FEP's wide-ranging explanatory power (Friston 2019b; Hohwy 2013; Kirchhoff et al. 2022). On the other side, there are antirealist accounts that voice reservations about FEP's lack of detailed elucidation concerning the mechanical details (Baltieri et al. 2020; Colombo & Palacios 2021; Klein 2018; van Es & Hipólito 2020). It is important to note that there is no definitive way to unequivocally support either realism or antirealism concerning the FEP. Each perspective has valid points in its favour. For instance, the FEP exhibits remarkable unifying power, as claimed by the realists. However, it lacks a comprehensive account of the underlying mechanisms, as argued by the antirealists. The structural realist perspective presented here aims to acknowledge the strengths of both the realist and antirealist stances on the FEP. These structures are unifying in the sense that they underlie the fundamental principles of thermodynamics, information theory, and other related fields. Thus, a best-of-both-worlds strategy in this context would comprise the realist emphasis on the underlying unificatory patterns[2] that underpin the maximal explanatory

[1] The reason for adopting this middle ground is that it does justice to evidence in favour of both realism and antirealism. Pure realist or antirealist stances, on the other hand, tend to overlook significant facts that support their counterpart. For instance, realists may overlook viable evidence in favour of theoretical change, while antirealists might ignore evidence supporting structural continuity.

[2] It's important to clarify that the concept of unification under discussion in the context of this paper and Structural Realism (SR) in general is not the traditional notion of unification in explanatory unification. Instead, it pertains to the structural continuity and formal commonality at the level of formalism that underlies diverse theories, fields or structures. This form of unification signifies theoretical progress. This point will be elaborated upon in the paper with reference to classical theories, such as Maxwell's equations and Fresnel's equations, in conjunction with the discussion of the FEP. This form of unification has also been employed at an ontological level to tackle metaphysical underdetermination (French 2011a), and Worrall (1989) has applied it in the context of addressing the question of scientific progress at the structural level underlying radical theoretical changes.

scope of FEP with a concession on the FEP's relative lack of elaboration on mechanical components of life and cognition.[3]

The first goal of this paper is to show that the viable philosophical stance on FEP must have the best points of realist and antirealist interpretations on its side. It acknowledges the existing gap that requires filling with mechanistic explanations while also advocating for realism concerning the structural relations that underlie FEP and unify it with the foundational equations of thermodynamics and information theory. SR about FEP brings about the best of both worlds. The claim of the continuity and unity of FEP with theories of thermodynamics and information lies at the centre of the tapestry of this paper. The paper aspires to characterise this continuity and unity in the same spirit that Worrall (1989) did characterise the progress of optics at the level of structure of, say, theories of optics and electrodynamics.

The paper is structured as follows: it begins by outlining the Free Energy Principle (FEP) and highlighting the differences between realist and antirealist interpretations of FEP. Subsequently, I examine Worrall's strategy for developing Structural Realism in the philosophy of science. I then demonstrate how his account of structural continuity, exemplified in the case of the history of optics encompassing Fresnel's equations and Maxwell's equations, can also be applied to FEP. This extends to the relevant structural continuity and integration in thermodynamics and information theory, albeit at the level of their formalism. I argue that Structural Realism concerning FEP can reconcile the valid aspects of both realism and antirealism regarding FEP.

2 The Free Energy Principle

The free energy principle (FEP) is at the centre of a unifying framework for characterising notions of cognition, life, and action (Friston 2009; 2012; 2019a). FEP's description of life and cognition is closely related to thermodynamics, where self-organisation is described in terms of the dynamics of dissipative structures under the second law of thermodynamics (Ueltzhöffer et al. 2021). More specifically, FEP explains life and cognition along the lines of adaptive behaviour under minimising an upper bound on surprise. This upper bound is articulated in terms of variational free energy, and surprise is the negative log-likelihood of sensory samples (Friston 2010; 2013). From this perspective, cognisant and living entities are self-organising systems that aim to reach the state of non-equilibrium steady states (for the relationship between cognition and life under FEP (Kirchhoff 2018). In general, according to FEP, ergodic dynamic random systems are capable of exhibiting life and cognition. When applied to life, FEP indicates that the organism maximises its survival by avoiding surprising states, which means that the organism tends to remain in non-equilibrium steady states. When applied to cognition, FEP indicates that the cognisant organism avoids surprising states by learning (via adaptive behaviour) to form internal cognitive models of the environment and updating its models of worldly structures by minimising their prediction errors through hierarchical cognitive structures (Friston and Stephan 2007). Because under ergodic assumptions, entropy is defined as the long-term average surprise, minimising free

[3] Mechanisms could be defined in a number of different ways. For the time being, let us defer to Glennan's definition: "A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interaction between parts can be characterized by direct, invariant, change-relating generalizations (Glennan 2002, 344)".

energy puts an upper bound on the entropy of the system (Friston 2013). Hence the relation between life, cognition and laws of physics (thermodynamics). Active inference under FEP consists of updating the internal states of the organism in a way that leads to Bayes-optimal perception or Bayes-optimal action (when the system engages in updating its active states).

In order to set a distinction between internal and external spaces (or active and sensory states), FEP uses Markov blankets, which are Bayesian network structures (Pearl 1988). A Markov blanket of a node consists of the node itself, its parents, its children and co-parents of its children. In the context of FEP, Markov blankets are used to compartmentalise the system by setting probabilistic independence between internal and external states, so as to explain how the active states affect the external states but remain unaffected by them, and how sensory states affect the internal states without being affected by them (Hipólito et al. 2021). Markov blankets apply at various scales, stretching from a population of neurons to brain regions, whole brains, whole organisms, and communities of organisms (Hesp et al. 2019).

On some occasions, the debate between realists and antirealists revolves around Markov blankets. Antirealists question the validity of making epistemological or ontological commitments to Markov blankets, sometimes referred to as Friston blankets in that context (Beni 2021; Bruineberg et al. 2021). On the other hand, realists advocate for the legitimacy of scientific realism concerning the Bayesian representation of cognitive processes. However, it's important to note that the reliance on Markov blankets as providing a viable representation of specific cognitive aspects doesn't imply that everything described by Bayesian networks is 'literally' true. For instance, asserting the presence of Markov blankets as actual anatomical components within the brain and cognitive systems is an oversimplification and does not need to be part of realist commitments (Kirchhoff et al. 2022; Kiverstein and Kirchhoff 2022). There is no indication that the realist and antirealist camps are making preparations for a truce in the foreseeable future. In the next section, I shall speak more on the discord between the realist and antirealist perspectives on FEP.

3 A Short Remark on Realist and Antirealist Readings of FEP

The modelling process is complicated, and it involves stages such as abstraction, generalisation, imputation and keying up (Frigg and Hartmann 2018; Frigg and Nguyen 2017). By extension, the issue of realism about the content of scientific models is complicated, mainly because there is no straightforward way of reading the features of scientific models into their target systems (Frigg 2010; Godfrey-Smith 2009). This paper is concerned with realism or antirealism about the class of scientific models that regiment the theoretical claims of FEP (I call them free energy models).

The antirealist regarding FEP assumes that, while free energy models might serve practical purposes, they would not offer a comprehensive account of cognitive mechanisms (see Colombo et al. 2021). Even proponents of the FEP may concede that the antirealist is correct in assuming that the FEP does not provide a detailed account of the implemented neuronal mechanisms of cognition (Beni 2022). This is why I believe that certain aspects of antirealism about the FEP must be taken seriously. On the other hand, there are also realist readings of FEP. A realist in this context can argue that realism about the free energy models provides the best explanation for the empirical adequacy of such models. For example, the empirical adequacy of the Bayesian models of the brain could be explained by

suggesting that the brain actually operates on the basis of minimizing the free energy of its models of the environment:

> [T]here is converging evidence that the brain is a Bayesian mechanism. This evidence comes from our conception of perception, from empirical studies of perception and cognition, from computational theory, from epistemology, and increasingly from neuroanatomy and neuroimaging. The best explanation of the occurrence of this evidence is that the brain is a Bayesian mechanism that is, in fact, engaged in inference, belief, and decision (Hohwy 2013, 25).

Also, FEP's unifying formalism may play a role in bolstering realism about FEP. I will address the question of why to take unifying formalism as grounds for realism in Sect. 6 6 of this paper. But for the time being, suffice it to say that not only in the philosophy of science but also in metaphysics unifying formalism sometimes (when it is apt) provides grounds for positing entities. For example, Jonathan Schaffer claims that when some phenomena are aptly formally unified, then there is a strong reason to assume that there is room for positing the unifier (Schaffer 2016, 153). On the same logic, the FEP's formal framework unifies various theoretical insights from thermodynamics, information theory, and machine learning. Therefore, the FEP can be considered a unifying principle.

FEP does indeed come with grand unificatory claims (Beni 2018; Friston 2010; Hohwy 2013, 96). Arguably the variational Bayesian of minimising prediction error under FEP accommodates and unifies streams of research "in machine learning, in psychophysics, in cognitive and computational neuroscience and (increasingly) in computational neuropsychiatry (Clark 2016, 294)". FEP's capacity to unify theories of the brain with information theory, computation, machine learning, and physics on the basis of its formalism can furnish new grounds for realism about FEP. To understand how unification can lead to realism in this context, it's crucial to note that the kind of unification being discussed here goes beyond unificatory explanations, which involve explaining several phenomena by unifying them. The central focus here is on unifying through the specification of structural commonalities across different disciplines and fields. In other words, unification highlights the formal consistency across thermodynamics, information theory and FEP, offering support for a nuanced form of realism that respects the valid antirealist concerns regarding the limited mechanistic details in FEP. To substantiate this claim and to illustrate how structural unification can underpin realism, the next section will draw upon examples from Worrall's version of SR.

4 Structural Realism: the Best of Both Worlds

Structural Realism (SR) has been something of official wisdom in the philosophy of science for the last couple of decades. Coming in various shapes and flavours, SR has its stronghold in the philosophy of physics (Esfeld and Lam 2008; French 2014; Ladyman and Ross 2007) (although there have also been attempts at projecting SR into the philosophy of special sciences (Beni 2019b; French 2011b; Hasselman et al. 2010; Ross 2008)). SR has precedents in the works of Bertrand Russell, Arthur Eddington and Henri Poincaré, but it has been articulated and introduced in its new guise in John Worrall's (1989) paper, 'Structural Realism: the best of both worlds?' As the title of Worrall's paper aptly indicates, SR aims to bring the best of both worlds together.

The two worlds that are mentioned in the title of Worrall's paper (as well as the present one) are the realist and antirealist worlds. Advocating the famous 'no miracles' argument, scholars of the realist world assume that the empirical success of science is due to the essential or approximate truth of theories, whereas denizens of the antirealist world draw attention to historical facts of radical theoretical shifts to argue that theories lack descriptive, truth-conducive (even approximately truth-conducive) content but only provide scaffolding for the experimental laws (Worrall 1989, 100). Interestingly enough, Worrall points out that although realist and antirealist positions pull in opposite directions, a satisfactory stance in the philosophy of science must have both arguments on its side. Building upon the historical works of Pierre Duhem and Henri Poincaré, Worrall argues that such an intermediary stance should be stated in terms of realism about the underlying structure of theories, rather than their theoretical content. To substantiate his point, Worrall draws attention to the complete formal continuity between Fresnel's and Maxwell's theories (Worrall 1989, 119). I shall shortly review this part of his paper.

This is Fresnel's equation:

$$R/I = \tan(i - r)/\tan(i + r)$$

$$R'/I' = \sin(i - r)/\sin(i + r)$$

$$X/I = (2sinr \cdot 2cosi)/(sin(i + r)\, cos(i - r))$$

$$X'/I' = (2sinr \cdot cosi)/sin(i + r)$$

In the context of Fresnel's equation, I^2, R^2, and X^2 are the intensities of polarised components in the place of incidence of reflected and refracted beams, and I'^2, R'^2, and X'^2 are components polarised at right angles, and r and i are the angles made by the refracted beams and planes. If we insist on construing the equations in terms of object-oriented ontology,[4] Fresnel's theory would describe light in terms of vibrations transmitted through a mechanical medium. But it is a well-known fact that Maxwell's theory renounces the idea of elastic ether as the vibrating medium, and as such Maxwell's theory confutes the ontological implications of Fresnel's theory. Despite this significant ontological disagreement, even from the perspective of Maxwell's theory, Fresnel's equations get the structure of electric and magnetic field strengths precisely enough. Maxwell's equation could preserve the structure of Fresnel's equations 'directly' and 'fully' by interpreting I, X, R, and other variables in terms of the amplitudes of the vibration of relevant electric vectors (Worrall 1989, 119). This means that, although the implications of Fresnel's theory about the nature of light were not correct, the theory still provides reliable theoretical insights into the structural properties of optical phenomena, which is quite enough for vouchsafing scientific progress at the structural level, despite conceding the possibility of radical replacements at the level of theoretical content. This proposal indeed combines the strength of the realist take on the relation between the empirical adequacy of theories and the veracity of their (*structural*) representations with the viability of the antirealist view on radical

[4] Object-oriented ontology adopts individual objects as primary subjects of ontological commitments, which is notably distinct from the perspective of structural realism (SR). SR posits that structures are the subjects of epistemological and/or ontological commitments.

theoretical changes. The result is SR, a moderate version of realism that is not saddled with presumptions about the nature of individual objects as the referents of theoretical terms.

It is important to emphasize what commitments of Worrall's statement of SR are not: SR does not assume that converging or continuous theories refer to the same physical entities. To the extent that the structural continuity of the theories of Fresnel and Maxwell is at issue, SR does not even need to assume the two theories are empirically equivalent. It can be readily observed that the scope of empirical application of Maxwell's theory is far wider than Fresnel's, and predictions of the two theories are not always the same. The philosophical commitments of SR are quite modest, in the sense that it is only committed to underlying commonality at the level of structure of theories. And Worrall's epistemic version of SR simply holds that science is in the business of providing reliable *structural knowledge* of the world, without going so far as to provide a watertight argument to the effect that the mind-independent world is populated by objects that are described by theories. Would the modesty of epistemic SR undermine its realist component? I believe the interpretation of the term 'realism' depends on one's perspective. Overall, Worrall's version of SR seeks to achieve a delicate equilibrium between complete realism and instrumentalism. It revolves around adopting a realist stance within the confines of one's means.

It might be worth noting that in the later sections of his paper, Worrall directs his attention towards demonstrating that the structural realist approach can also be applied to address versions of antirealism rooted in the metaphysical implications of quantum mechanics (assuming that modern physics is not aligned with object-oriented metaphysics). This notion has been further developed into a robust ontic structuralist version of SR (French 2011a; French and Ladyman 2003; Ladyman 1998). While I won't delve deeply into the discussion of the ontic version of SR in this paper, this observation hints at some convergence between the epistemic and ontic versions of SR.

Be that as it may, recently, there have been quite a few attempts at challenging the realist interpretation of FEP (Colombo et al. 2021; Colombo and Palacios 2021; van Es and Hipólito 2020). In the subsequent sections of the paper, I will modify the structural realism (and the best of both worlds strategy) to counteract the challenges posed by antirealists regarding FEP. However, this will be done without veering into ontological or even epistemological commitments concerning the referents of FEP's theoretical terms (as individual objects). This implies that rather than attempting to address antirealist interpretations of FEP on an individual and direct basis, I will concentrate on demonstrating how the progression of FEP from thermodynamics and information theory maintains sufficient structural continuity. This continuity justifies a modest form of structural realism about FEP, even though the theoretical content of these theories differs significantly.

5 Tracing the Structural Continuity

In the previous section, I briefly outlined Worrall's articulation of SR and his account of structural continuity across theoretical changes. Before that, I set a distinction between two types of grounds for realism about FEP, with the latter set of grounds mainly consisting of the consideration of the unificatory formalism of FEP. In this section, I put flesh on the skeletal scheme of this structural swath of realism about FEP, by elaborating on the structural continuity across thermodynamics, information theory and FEP. I argue that FEP's reliance on the central concept of entropy can be seen as a form of structural progress originating from thermodynamics and information theory, even though it does

not adhere to the same theoretical commitments as these fields (if we insist on interpreting theoretical commitments in terms of orthodox ontology). Characterising the said formal continuity underneath the theoretical evolution of the FEP will substantiate a viable structural realist account of FEP (or rather about the formal structures that lie beneath thermodynamics, information theory and FEP). This structural realist reading will have the best of both realist and antirealist interpretations of FEP on its side.

SR has been already nominated as the philosophical confederate of FEP on a couple of occasions. The fleeting remark on the amicable fellowship of SR and FEP has been based on insights into the structural nature of integrity that the organism strives to achieve under FEP (Friston et al. 2020; Wiese and Friston 2021). Generally, the unity that FEP is supposed to bring to otherwise diversified studies of cognition, action, etc., is mainly grounded in the underlying patterns that lie beneath the disciplinary compartmentalisations (Beni 2019a, b, Ch. 5). The underlying pattern could be neatly articulated in terms of Friston's account in an informational geometric framework that regiments the dynamics of FEP. The general assumption, which speaks to SR directly, holds that in "different biological systems, free energy is minimised in dynamically equivalent, but mechanistically heterogeneous, ways (Ramstead et al. 2017, 8)". Thus, here are already voices (or perhaps whispers) that appreciate a structural realist interpretation of FEP. However, to provide stronger support for the structural realist interpretation of FEP, I embark on reconstructing Worrall's best-of-both-worlds strategy by specifying structural continuity across the historical evolution of FEP from thermodynamics and information theory. As we have seen before in this paper, there is already a disparity between the realist and anti-realist interpretations of FEP, with realists building their claim upon explanatory power and unificatory scope of FEP (Clark 2016; Hohwy 2020) and antirealists emphasising the lack of enough theoretical elaboration on the nature of local mechanisms that enable minimisation of free energy in living or cognitive systems (Colombo and Palacios 2021; Colombo and Wright 2018). The situation—namely the disparity between realist and antirealist stances on FEP—provides a nice occasion for developing a structural realist stance on FEP, more or less in the same way that a similar situation in the philosophy of physics motivated Worrall's best of both worlds strategy. On both occasions (the present study and Worrall's original proposal) SR evolves into an intermediary stance that retains the worthwhile insights of both parties by identifying structural continuity underneath theoretical diversities. In this regard, the loci of (epistemic) realist commitments remain centred on the mathematical structures of the FEP. These structures serve as the binding elements that bridge the fundamental insights of thermodynamics and information theory. This approach allows SR to uphold its intermediary stance, safeguarding the valuable contributions from both sides by uncovering structural continuity amid theoretical diversities.

We continue our technical exploration of FEP, picking up where we left off with our discussion of the affinity between FEP and Gibbs energy. We examine the situation through the lens of Boltzmann's equation. The fundamental notion of entropy has been introduced by Ludwig Boltzmann in the 1870s in terms of a basic formula of thermodynamics:

$$S = k_B log W$$

In this equation, k_B is Boltzmann constant, S denotes entropy, and W denotes the number of microstates that correspond to a given microstate of a given thermodynamic system. In Boltzmann's equation, it is generally assumed that microstates have equal

probabilities, but when the probability of microstates is not equal, Boltzmann's equation evolves to the articulation of Gibbs entropy, which is:

$$S_G = -k \sum_{i=1}^{k} p\,(i)\, logp\,(i)$$

In this equation, S_G denotes entropy, k is a positive constant, and p(i) denotes the probability of the microstates of the system in state i.

It has been correctly pointed out that the agreement between Boltzmannian and Gibbsian formulations is not universal, and they are not even empirically equivalent (Frigg and Werndl 2019). More fundamentally, the characterisation of physical processes and events is not quite similar across Boltzmann's and Gibb's equations. To establish their point, Frigg and Wendl argue that Gibbs' equation is unconcerned with the notions of motion or dynamical laws of the evolution of systems, whereas dynamical considerations have a central status in Boltzmann's equation (Frigg and Werndl 2019, 430–431). The point is well-taken. But if we take a structural realist stance on the situation we would not need to make commitments to the physical characterisation of equations. SR is only committed to the structural commonality between the mentioned equations, despite appreciating the divergence in their characterisation of physical entities. (In the next section, I will refer to Ladyman and Ross (2007) to elaborate on how the unifying formalism across accounts of work and energy in the works of Helmholtz and Joule has motivated realism about structures).

The above-mentioned point, about the diversity of physical characterisations, becomes even more obvious if we compare the thermodynamic and information-theoretic notions of entropy. The physical character of concepts of theories (or the physical referents) are different across thermodynamics and information theory (one of them refers to the distribution of thermodynamic states, whereas the other refers to message states), and at the same time, there is remarkable formal convergence between them. This is because the form of Gibb's equation is fully and completely similar to Claude Shannon's (1948) articulation of the notion of uncertainty or entropy in the context of information theory:

$$H = -\sum_{i=1}^{n} P\,(x_i)\, logP\,(x_i)$$

The formal convergence can be substantiated in various ways. So, for example, if we interpret the units of Shannon's equation in terms of the Boltzmann constant, we would end up with Gibb's equation. Indeed, the resemblance (between the two statements of entropy across thermodynamics and information theory) had been brought to Shannon's attention as early as the 1930s by von Neumann (Shannon and Weaver 1949, 3). So, there is a clear line of formal unity and continuity across thermodynamics and information theory.[5] Now,

[5] It may be true that the epistemic (and possibly ontic) affinity between the thermodynamics and information theory could not be established solely based on the formal convergence at the level of mathematical syntax (for conceptual elaborations will be necessary as well) (Jaynes 1957). However, to the extent that Worrall's statement of SR is at issue, the epistemic and ontic implications of theoretical terms per se are not important (at least not as much as the *structural convergence* of the form of theories could be important). Moreover, Jaynes's own account of the conceptual link between information theory and thermodynamics, and the relevant notions, such as ergodicity, etc., is in practice constructed. That is to say, Jaynes's own account of conceptual convergence between information theory and thermodynamics builds upon *formal facts* grounded in structural continuity across information theory and thermodynamics. The said facts concern the "maximization of entropy", and Jaynes mentiones the maximisation of specific probability distributions via entropy subject to certain constraints as an example of such mathematical facts (Jaynes 1957,

our study concerns realism about FEP. Despite the diversity of the physical concepts, there is formal continuity across thermodynamics and information theory on the one hand and FEP on the other hand. I shall flesh out this claim in the remainder of the paper.

Shannon defines ergodicity in terms of statistical regularity or homogeneity, assuming that in an ergodic process, "every sequence produced by the process is the same in statistical properties (Shannon and Weaver 1949, 45)". The notion of ergodicity has a central role in Friston's statement of FEP and its account of the regularity (or high probability) of the organism's tendency to avoid the dissipation of its states over the phase space. To survive, organisms occupy a bounded set of states amongst a total set of possible states that they can occupy, assuming that "any ergodic random dynamical system that possesses a Markov blanket will appear to actively maintain its structural and dynamical integrity" (Friston 2013, 2)". The fundamental notion—that can as well be spelt out structurally—is 'entropy'. That is to say, FEP's notion of ergodicity is grounded in entropy, which is structurally specified across fields of information theory and thermodynamics; the high probability of the organism occupying the same states under the assumption of ergodicity speaks to the low entropy of probability distribution of the places that the organism occupies in the phase space. The entropy is bounded by the free energy which serves as an upper bound on surprise (whose time average is entropy). At the theoretical level (i.e., level of theoretical biology), the organism's tendency to avoid states with high entropy is explained in evolutionary terms, by saying that the organisms that succeeded in avoiding entropic states have been selected over the organisms that failed to do so (Ramstead et al. 2017, 2). This biological insight has been constricted into the neat formalism of FEP, which defines free energy in terms of relative entropy represented by the Kullback-Liebler divergence and surprise: In general, assuming that $\psi \in \Psi$ represents the hidden state of the world that causes the sensory state represented by $s \in S$, and $\lambda \in \Lambda$ represents the internal state of the system, then:

$$F = Energy - Entropy = -\langle lnp(s|\psi\rangle_q + lnq(\psi)_q$$

This takes us back to Gibbs free energy, which in simplest form is stated as

$$\Delta G = \Delta H - T\Delta S$$

where T is the temperature, S is the entropy, and H is defined as the sum of the internal energy and the product of pressure and volume (U + pV). To emphasise my point once more, there is a theoretical shift across Gibbs's equation and the FEP equation. FEP bounds the notion of Gibbs's free energy and applies it to a specific context of life and cognition—where structures accommodate notions of the sensory states of an agent or its organism and its environment, rather than the ingredients of Gibbs's theory, such as internal energy, pressure, or volume. However, to the extent that FEP enacts the formal notion of entropy, the structural continuity between Gibbs's theory and Friston's theory endures. Below, I briefly show how FEP delimits and applies Gibbs's thermodynamic notion to a specific context of life.

Footnote 5 (continued)
621)). Ergodicity is another common notion that has been considered to be a mathematical fact in this context. The relational articulation of the concept of entropy by drawing on relevant mathematical-structural facts in the way that is pioneered by Jaynes is in complete harmony with the foundational insights of SR.

For any given instance of Gibbs energy $G(\psi, s, a, \lambda) = -lnp(\psi, s, a, \lambda)$, there is a free energy $F(s, a, \lambda)$ that describes the flow of internal and active states:

$$f_{\lambda}(s, a, \lambda) = -(\Gamma + R) \cdot \nabla_{\lambda} F$$
$$f_{a}(s, a, \lambda) = -(\Gamma + R) \cdot \nabla_{a} F$$
$$AndF(s, a, \lambda) = \int_{\Psi} q(\psi|\lambda) \, ln \, \frac{p(\psi, s, a, \lambda|m)}{q(\psi|\lambda)} \, d\psi = Eq\,[G(\psi, s, a, \lambda) - Hq(\psi|\mu)]$$

where ψ, s, a, λ respectively denote specific external, sensory, active and internal states of the system, F is Friston's free energy, G denotes Gibbs energy, and Γ is a diffusion tensor. According to Friston (2013, 4), in the context of FEP, $p(\psi, s, a, \lambda)$ defines the ergodic density, which is the probability density function over the said states of the organism-environment dynamical relationship, and $q(\psi|\lambda)$ is variational density, which is an arbitrary function over external states parametrised by internal states. The last line of the equation once more indicates that free energy can be spelt out in terms of the expected Gibbs energy minus the entropy of the variational density. FEP draws on the notion of entropy, which underlies structural continuity between thermodynamics, information theory and FEP. However, by construing variables s and a as sensory and active states and ψ *and* λ as external and internal states of an organism, FEP narrows down the wider scope of theories of thermodynamics and instantiates them into the context of life and cognition. The uniformity of the form or structure across the shift from thermodynamics to information theory and eventually FEP provides leverage for a structural realist stance on the common underlying structure of FEP, albeit without providing support for realism about the theoretical content. As such, I reconstructed Worrall's example of formal continuity across Fresnel's and Maxwell's respective theories.

This section elaborated on the structural continuity across thermodynamics, information theory and FEP. All that said, the lingering question is why suppose that the mathematical convergence across the sets of equations of fields of thermodynamics, information theory, and FEP grounds *realism*. I shall discuss this issue fully in the next section.

6 Discussion

Does the unifying formalism ground (structural) realism about free energy models though? To a first approximation, (orthodox) realism is supposed to be centred on a correspondence between theories and mind-independent reality. Realists, it might be assumed, must trade in the currency of truth and objective facts, rather than formal unification. The problem with this orthodox version of realism is that, alas, we do not have mind-independent (or theory-independent) access to the world in itself. This means that we must rely on our cognitive faculties as well as the resources of our scientific knowledge to be able to speak meaningfully about reality. Our conception of reality is reliant on scientific knowledge. Unifying formalism underwrites the progress of scientific knowledge, which shapes our conception of reality. There is ample historical evidence for this claim. For example, as Ladyman and Ross (2007, 42) argue, Faraday's research on electromagnetic induction unifies different concepts of electrostatic or induced electricity. Also, progress in theories of work and energy in physics manifested with a great unifying trajectory, e.g., there is a commonality in formalism across Helmholtz' formalisation of the conservation of the sum of kinetic and potential energy in rational mechanics and Joule's account of the relation between heat and mechanical work (Ladyman and Ross 2007, 42). Even though unifying formalism across

theories does not lead to watertight metaphysical proof for realism, it provides pragmatic grounds for optimism about scientific progress.

A full-blown defence of naturalism remains beyond the scope of this paper, but to the extent that we concede the reliance on science and its history for fleshing out philosophical stances, the same unifying trajectory that warrants (structural) realism about the theoretical implications of modern physics (Ladyman & Ross 2007, 71, 142, 190 ff.) can also support the claim of FEP to structural realism on grounds of its strong unifying links with thermodynamics and information theory. In this context, it is important to underscore the historical and formal evolution that has transpired. Thermodynamics, which laid the groundwork for understanding energy and entropy, ultimately led to the development of information theory. This progression marked a transition from the physical domain to the realm of information and cognition. The FEP emerges as the latest chapter in this evolution, unifying these domains with its formalism. Through the lens of the FEP, the structural continuity from thermodynamics to information theory and, subsequently, to cognitive science is evident. This continuity is not just a matter of historical development but also one of formal alignment. The FEP's formalism harmonizes and builds upon the principles of thermodynamics and information theory, uniting them in its account of life, cognition, and adaptive behaviour. Consequently, the FEP's journey represents an intricate interplay of historical progression and formal integration, making it a compelling case for structural realism within the philosophy of science. The general insight here is that while the notions of entropy and ergodicity retain their structural integrity across the fields of thermodynamics, information theory and FEP, the theoretical interpretation of these elements varies across different fields. Take the following tuple: (Ω, Ψ, S, A, Λ). Let Ω denote the non-empty set of the system's state space. In the context of thermodynamics, the tuple represents the elements of the evolution of the system over time, such as the space of micro-constituents, the evolution functions and so on. In the context of the information theory, these elements should be rephrased to represent the modes of transfer of information from the sender to a receiver and the effect of the environment or information channel. The respective theoretical interpretations do not remain the same across thermodynamics, information theory and FEP. FEP renormalises the thermodynamic (and information-theoretic) concepts by applying them to specific contexts in biology and the cognitive sciences, in terms of agents, their sensory and active states, and their dynamical relationship with the environment. As such, there is no room for adopting a realist perspective concerning the theoretical content encapsulated in FEP's account of cognition and action, let alone embracing realism regarding FEP's description of the implemented cognitive mechanisms. To illustrate this, consider the context of thermodynamics, where, for instance, the equation $S_G = -k \sum_{i=1}^{k} p\,(i)\, logp\,(i)$. In this equation, $p\ (i)$ represents the probabilities of microstates corresponding to specific thermodynamic states. More precisely, within the framework of Boltzmann's equation, 'W' symbolizes the count of microstates that correspond to various macroscopic thermodynamic states.[6] However, in the context of information theory, the same piece of mathematics represents the probability of selecting a message from a message state within any given message space. On the other hand, within the framework of the Free Energy Principle (FEP), this distribution of probabilities takes on a distinct interpretation. It pertains to sensory perception, adaptive actions and behaviour. Additionally, it extends to the organism's self-information and prediction error, all of which arise from the dynamic interplay between the organism's predictions and actual events. This means that

[6] As I remarked earlier in this section, there is even some divergence in theoretical characterisation of entities in the diverse contexts of Boltzmann's equation and Gibbs's equation.

Friston's articulation of the FEP capitalizes on the structural parallels between information theory and thermodynamics, offering a unified and rigorous framework for understanding life, cognition and adaptive behaviour. This continuity from thermodynamics to information theory and, ultimately, the FEP underscores the sense of scientific progress and continuity that has been the unifying concept of SR.

7 Concluding Remarks

This paper applies John Worrall's (1989) strategy of Structural Realism (SR) to the Free Energy Principle (FEP) and engages with the ongoing debate concerning realist and antirealist interpretations of the FEP. The objective is to establish a philosophical position on the FEP that effectively integrates the valid aspects of both realism and antirealism. Worrall's SR offers a middle ground, emphasizing the significance of continuity and progress at the level of formal structures rather than delving into the local mechanisms that might underlie the theoretical terms within cognitive science theories. This paper focused on structural commonalities shared among the FEP, thermodynamics, and information theory, providing a robust foundation for a structural realist interpretation of the FEP. By accentuating structural continuity and unity across these domains, this paper contends that the FEP justifies a modest form of structural realism that can reconcile the valuable insights of both realism and antirealism.

Acknowledgements The debt to anonymous referees of this journal is gratefully acknowledged.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Declarations

Conflict of interest The author, Majid D. Beni, declares that he has no conflict of interest.

Ethical Approval This article does not contain any studies with animals performed by any of the authors. This article does not contain any studies with human participants or animals performed by any of the authors.

References

Baltieri, M., C.L. Buckley, and J. Bruineberg. 2020. Predictions in the eye of the beholder: An active inference account of Watt governors. *Artificial Life Conference Proceedings 32*, 121–129.

Beni, M.D. 2018. 'The Reward of Unification: A Realist Reading of the Predictive Processing Theory': *New Ideas in Psychology*. https://doi.org/10.1016/j.newideapsych.2017.10.001.

Beni, M.D. 2019a. 'Conjuring Cognitive Structures: Towards a Unified Model of Cognition': *Model-Based Reasoning in Science and Technology*. https://doi.org/10.1007/978-3-030-32722-4_10.

Beni, M.D. 2019b. *Structuring the Self*. Palgrave Macmillan.

Beni, M.D. 2021. 'A Critical Analysis of Markovian Monism': *Synthese*. https://doi.org/10.1007/S11229-021-03075-X.
Beni, M.D. 2022. 'Dosis Sola Facit Venenum: Reconceptualising Biological Realism': *Biology & Philosophy*. https://doi.org/10.1007/S10539-022-09884-9.
Bruineberg, J., K. Dolega, J. Dewhurst, and M. Baltieri. 2021. 'The Emperor's New Markov Blankets': *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X21002351.
Clark, A. 2016. 'Surfing Uncertainty': Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190217013.001.0001.
Clark, A. 2020. 'Beyond Desire? Agency, Choice, and the Predictive Mind': *Australasian Journal of Philosophy*. https://doi.org/10.1080/00048402.2019.1602661.
Colombo, M., L. Elkin, and S. Hartmann. 2021. 'Being Realist About Bayes, and the Predictive Processing Theory of Mind': *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axy059.
Colombo, M., and P. Palacios. 2021. 'Non-equilibrium Thermodynamics and the Free Energy Principle in Biology': *Biology & Philosophy*. https://doi.org/10.1007/S10539-021-09818-X.
Colombo, M., and C. Wright. 2018. 'First Principles in the Life Sciences: The Free-Energy Principle, Organicism, and Mechanism': *Synthese*. https://doi.org/10.1007/S11229-018-01932-W.
Esfeld, M., and V. Lam. 2008. 'Moderate Structural Realism about Space-Time': *Synthese*. https://doi.org/10.1007/s11229-006-9076-2.
French, S. 2006. 'VI—Structure as a Weapon of the Realist': *Proceedings of the Aristotelian Society (hardback)*. https://doi.org/10.1111/j.1467-9264.2006.00143.x.
French, S. 2011a. 'Metaphysical Underdetermination: Why Worry?': *Synthese*. https://doi.org/10.1007/s11229-009-9598-5.
French, S. 2011b. 'Shifting to Structures in Physics and Biology: A Prophylactic for Promiscuous Realism': *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*. https://doi.org/10.1016/J.SHPSC.2010.11.023.
French, S. 2014. 'The Structure of the World: Metaphysics and Representation': *Oxford University Press*. https://doi.org/10.1093/acprof:oso/9780199684847.001.0001.
French, S., and J. Ladyman. 2003. 'Remodelling Structural Realism: Quantum physics and the Metaphysics of Structure': *Synthese*. https://doi.org/10.1023/A:1024156116636.
Frigg, R. 2010. 'Models and Fiction': *Synthese*. https://doi.org/10.1007/s11229-009-9505-0.
Frigg, R., and Hartmann, S. 2018. Models in Science. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, (Spring 2020): https://plato.stanford.edu/archives/spr2020/entries/models-science/.
Frigg, R., and J. Nguyen. 2017. 'Models and Representation': In *Springer Handbook*, 49–102. https://doi.org/10.1007/978-3-319-30526-4_3.
Frigg, R., and C. Werndl. 2019. 'Statistical Mechanics: A Tale of Two Theories': *The Monist*. https://doi.org/10.1093/MONIST/ONZ018.
Friston, K.J. 1994. 'Functional and Effective Connectivity in Neuroimaging: A Synthesis': *Human Brain Mapping*. https://doi.org/10.1002/hbm.460020107.
Friston, K.J. 2009. 'The Free-Energy Principle: A Rough Guide to the Brain?': *Trends in Cognitive Sciences*. 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. https://doi.org/10.1038/nrn2787.
Friston, K.J. 2012. 'A Free Energy Principle for Biological Systems': *Entropy (Basel, Switzerland)*. 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. 2019a. A Free Energy Principle for a Particular Physics. arXiv reprint arXiv:1906.10184 (2019).
Friston, K.J. 2019b. 'Beyond the Desert Landscape': *Andy Clark and His Critics*. https://doi.org/10.1093/oso/9780190662813.003.0014.
Friston, K.J., and K.E. Stephan. 2007. 'Free-Energy and the Brain': *Synthese*. https://doi.org/10.1007/s11229-007-9237-y.
Friston, K.J., W. Wiese, and J.A. Hobson. 2020. 'Sentience and the Origins of Consciousness: From Cartesian Duality to Markovian Monism': *Entropy*. https://doi.org/10.3390/E22050516.
Glennan, S. 2002. 'Rethinking Mechanistic Explanation': *Philosophy of Science*. https://doi.org/10.1086/341857.
Godfrey-Smith, P. 2009. 'Models and Fictions in Science': *Philosophical Studies*. https://doi.org/10.1007/s11098-008-9313-2.

Hasselman, F., M.P. Seevinck, and R.F.A. Cox. 2010. 'Caught in the Undertow: There is Structure Beneath the Ontic Stream': *SSRN*. https://doi.org/10.2139/ssrn.2553223.
Hesp, C., M. Ramstead, K. Friston et al. 2019. 'A Multi-scale View of the Emergent Complexity of Life: A Free-Energy Proposal': *Springer Proceedings in Complexity*. https://doi.org/10.1007/978-3-030-00075-2_7.
Hipólito, I., M.J.D. Ramstead, T. Parr et al. 2021. 'Markov Blankets in the Brain': *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2021.02.003.
Hohwy, J. 2013. 'The Predictive Mind': *Oxford Academic*. https://doi.org/10.1093/acprof:oso/9780199682737.001.0001.
Hohwy, J. 2020. 'Self-supervision, Normativity and the Free Energy Principle': *Synthese*. https://doi.org/10.1007/s11229-020-02622-2.
Jaynes, E.T. 1957. 'Information Theory and Statistical Mechanics': *Physical Review*. https://doi.org/10.1103/PhysRev.106.620.
Kirchhoff, M. 2018. 'Hierarchical Markov Blankets and Adaptive Active Inference: 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.2017.12.009.
Kirchhoff, M.D., J. Kiverstein, and I. Robertson. 2022. 'The Literalist Fallacy and the Free Energy Principle: Model-Building, Scientific Realism, and Instrumentalism': *British Journal for the Philosophy of Science*. https://doi.org/10.1086/720861.
Kiverstein, J., and M. Kirchhoff. 2022. Scientific Realism About Friston Blankets without Literalism. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X22000267.
Klein, C. 2018. 'What Do Predictive Coders Want?': *Synthese*. 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*. https://doi.org/10.1016/S0039-3681(98)80129-5.
Ladyman, J. 2007. 'On the Identity and Diversity of Objects in a Structure': *Proceedings of the Aristotelian Society, Supplementary Volumes*. https://doi.org/10.1111/j.1467-8349.2007.00149.x.
Ladyman, J., and D. Ross. 2007. 'Every Thing Must Go': *Oxford Academic*. https://doi.org/10.1093/acprof:oso/9780199276196.001.0001.
Palm, G., T. Wennekers, N. Kogo, and C. Trengove. 2015. 'Is Predictive Coding Theory Articulated Enough to be Testable?': *Frontiers in Computational Neuroscience*. https://doi.org/10.3389/fncom.2015.00111.
Pearl, J. 1988. 'Probabilistic Reasoning in Intelligent Systems': The Morgan Kaufmann Series in Representation and Reasoning. https://doi.org/10.1016/c2009-0-27609-4.
Ramstead, M.J.D., P.B. Badcock, and K.J. Friston. 2017. 'Answering Schrödinger's Question: A Free-Energy Formulation': *Physics of Life Reviews*. https://doi.org/10.1016/J.PLREV.2017.09.001.
Ross, D. 2008. 'Ontic Structural Realism and Economics': *Philosophy of Science*. https://doi.org/10.1086/594518.
Schaffer, J. 2016. 'Ground Rules: Lessons from Wilson': *Scientific Composition and Metaphysical Ground*. https://doi.org/10.1057/978-1-137-56216-6_6
Shannon, C.E. 1948. 'A Mathematical Theory of Communication': *Bell System Technical Journal*. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.
Shannon, C.E., and W. Weaver. 1949. *The Mathematical Theory of Communication*. Illinois: University of Illinois Press.
Ueltzhöffer, K., L. Da Costa, D. Cialfi, and K. Friston. 2021. 'A Drive Towards Thermodynamic Efficiency for Dissipative Structures in Chemical Reaction Networks': *Entropy*. https://doi.org/10.3390/E23091115.
Van Es, T., and I. Hipólito. 2020. *Free-Energy Principle, Computationalism and Realism: A Tragedy*. https://philsci-archive.pitt.edu/18497/.
Wiese, W., and K.J. Friston. 2021. 'Examining the Continuity between Life and Mind: Is There a Continuity between Autopoietic Intentionality and Representationality?': *Philosophies*. https://doi.org/10.3390/PHILOSOPHIES6010018.
Worrall, J. 1989. 'Structural Realism: The Best of Both Worlds?': *Dialectica*. https://doi.org/10.1111/j.1746-8361.1989.tb00933.x.