Synthese (2024) 204:11
https://doi.org/10.1007/s11229-024-04660-6

ORIGINAL RESEARCH

Carving teleology at its joints

Majid D. Beni[1] · **Karl Friston**[2,3]

Received: 4 December 2023 / Accepted: 29 May 2024 / Published online: 26 June 2024
© The Author(s) 2024

Abstract
This paper addresses the conceptualisation and measurement of goal-directedness. Drawing inspiration from Ernst Mayr's demarcation between multiple meanings of teleology, we propose a refined approach that delineates different kinds of teleology/teleonomy based on the temporal depth of generative models of self-organising systems that evince free energy minimisation.

Keywords Free energy principle · Goal-directedness · Agency · Teleonomy · Teleology

1 Introduction

It is desirable to articulate the goal-oriented nature of (biotic) behaviour without reverting to vitalistic or anthropomorphic constructs. The impetus to account for the goal-directedness of biological systems and processes—unburdened by the constraints of scholastic teleology—gave rise to the discourse of teleonomy (Huxley, 1960; Mayr, 1998; Pittendrigh, 1958). This discourse has recently regained attention (Auletta, 2011; Dresow & Love, 2023; Gontier, 2022; Vane-Wright, 2022). This paper revisits Ernst Mayr's conceptualisation of the multiple meanings of teleology/teleon-

✉ Majid D. Beni
mbeni@metu.edu.tr

Karl Friston
k.friston@ucl.ac.uk

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

2 Queen Square Institute of Neurology, University College London, London WC1N 3AR, UK

3 VERSES AI Research Lab, Los Angeles, CA, USA

omy[1], as well as his account of teleonomy in terms of programmability (Mayr, 1985, 1988). Inspired by Mayr's insights, this paper seeks to reignite the project of delineating different types of teleology and integrating them with the concept of agency. We articulate the operational measure for specifying kinds of goal-directedness in terms of the 'temporal thickness' (Chouraqui, 2011) of generative models under the Free Energy Principle (FEP). The thickness or depth of models is associated with the ability to infer the past or future. We connect this to the notion of agency. While terms like 'agency' and 'teleology' often describe various levels of goal-directedness, our focus is to employ the criterion of temporal thickness in generative models under the FEP. We aim to differentiate between advanced forms of agentic goal-directedness—which is usually associated with the exchange of sentient organisms with their environment—and more rudimentary forms of goal-directedness in physical and biological domains. In this vein, our enterprise finesses preceding scientific and philosophical attempts to explicate goal-directedness on the basis of negative feedback to attain homeostasis with the environment (Nagel, 1977; Wiener, 1948). The FEP aligns with this tradition as it builds upon fundamental cybernetic concepts such as homeostasis and predictive control (Seth, 2014). However, inspired by Mayr's ideas, we take a further step to find an objective criterion for setting distinctions between kinds of goal-directedness.

Mayr's proposal takes on a renewed prescience within the realm of neuro-computational communication, particularly communications within the complex networks of bioelectric signalling that operate at the cellular level. These networks can be interpreted as possessing goal-directedness and agency, which are also manifested across multiple levels of organisational complexity (Blackiston et al., 2021, 2022; Clawson & Levin, 2022; Davies & Levin, 2023; Levin et al., 2020). Levin's research on the ensuing basal cognition posits bioelectric manipulability as a medium of teleonomy. His proposal blurs the boundaries between traditionally recognised purposeful agents (such as organisms shaped by natural selection) and sub-personal components, such as cells, which actively sample evidence for their existence. (Fields, Friston, Glazebrook, Levin et al., 2022; Kuchling et al., 2020).

This understanding of purpose arises from a multi-scale conception of collective intelligence, which consequently renders teleology/teleonomy a multi-scale aspect. This implies that intelligence emerges at various degrees of complexity across diverse levels of hierarchical information processing and does not require a characteristic scale. We aim to bridge Mayr's and Levin et al.'s proposals by developing a criterion that encompasses both perspectives and improves them. While Mayr's conceptualisation of teleonomy—in terms of programmability—provides a fine foundation, it requires further refinement to establish a comprehensive criterion for measuring teleology/teleonomy. On the other hand, in the context of Levin's work, the monolithic conception of goal-directedness and ensuing agency—across various levels—could be supplemented with a criterion for identifying the different kinds of teleology and agency across diverse scales. We specify this criterion in terms of the temporal depth

[1] This paper is not specifically concerned with the difference between 'teleonomy' and 'teleology'. While we are aware of the significance of that difference, unless specified otherwise, we use these notions interchangeably in this paper, leaving the discussion of the distinction between the two for later consideration.

or thickness of generative models under the FEP — to discern different types of teleonomy and agency across multiple levels. Within the FEP framework, generative models are probabilistic predictive mappings from causes (e.g., latent or hidden states of the world) to consequences. The thickness or depth of generative models bestows upon organisms the ability to infer past and future events. This implies that thick or deep generative models underwrite a sense of purposiveness and agency under the rubric of the Free Energy Principle (FEP). The ability to anticipate the future and model the (Markovian) world is crucial, given that the future depends upon the action of the system in question. This endows the system with a minimal kind of agency—namely, a system that acts upon the world in a future-pointing fashion. Our enterprise is, on the one hand, in line with the common practice in life science, aiming to classify and categorise entities and processes objectively. On the other hand—because we invoke a formal (computational) framework, namely the Free Energy Principle (FEP)—our enterprise aligns with Mayr's proposal to naturalise teleonomy, in terms of computation and programmability; read as Bayesian mechanics (Ramstead et al., 2023). In summary, we try to reinforce Mayr's distinction between purposeful (agentic) and historical (Markovian) teleologic processes using the free energy principle. Markovian processes in this context refer to processes that can be modelled in terms of Bayesian inference assuming conditional independence between internal and external states. The main point here is that not all processes that can be described in this fashion possess agentic goal-directedness. There are historical processes such as evolution that can be described as minimising free energy and thereby modelled in Bayesian terms without having human-like agentic teleology. Specifically, we identify agentic teleology with certain (conservative) systems that (appear to) engage in planning as inference, as distinct from historical processes that simply pursue paths of least action. This enables us to characterise certain kinds of teleologic processes as being equipped with generative models that have temporal depth (e.g., the anticipatory behaviour of animals), as opposed to (historical), Markovian self-organising processes (e.g., the weather or evolution). This attempt also pursues cybernetic efforts to naturalise purposiveness and teleology by providing a formal distinction between kinds of goal-directedness.

The paper is structured as follows. We commence by providing a concise overview of Mayr's characterisation of teleonomy in terms of programmability, we then consider the distribution of teleology/teleonomy across various scales of self-organisation. Then we briskly explicate the philosophical motivation driving our endeavour: to provide a measure of the degree of goal-directedness across diverse scales of self-organisation. After that, we unravel the centrepiece of our proposal: namely, that the Free Energy Principle (FEP) underwrites the difference between kinds of goal-directedness, on the basis of (temporal) thickness or depth of the generative models, for which evidence is sought.

2 Goal-directedness and computation

Ernst Mayr, an eminent evolutionary biologist, made significant contributions to the teleonomy discourse—in response to the seminal works of (Huxley, 1960; Pittendrigh, 1958)—by suggesting the goal-directedness of behaviour (and development) can be identified by examining the execution of implemented programs or codes of information processing in systems that exhibit goal-directed behaviour (Mayr, 1961, 1998, pp. 36–37). His proposition posits that teleonomy is contingent upon the functioning of a program, which may manifest as an evolutionary by-product, such as a genetic code, or an artificially created entity.

Interestingly, when introducing his concept—to specify teleonomy in terms of computation—Mayr also addressed the question of when it is appropriate to attribute purposiveness to target systems and processes and when it is not. He differentiated those systems that can justifiably be associated with purposiveness and *agency*, from *historical processes*—such as evolution—that cannot be specified as purposive entities. According to him:

> Where, then, is it legitimate to speak of purpose and purposiveness in nature, and where is it not? To this question we can now give a firm and unambiguous answer. An individual who–to use the language of the computer–has been "programmed" can act purposefully. Historical processes, however, cannot act purposefully. A bird that starts its migration, an insect that selects its host plant, an animal that avoids a predator, a male that displays to a female–they all act purposefully because they have been programmed to do so. (Mayr, 1961, pp. 1503–1504)

Not only does Mayr establish a clear distinction between teleonomic and non-teleonomic phenomena, but he (1998) also expands upon the diverse meanings of teleonomy. His substantial contribution serves as the foundation for our motivation to explore distinct kinds of teleonomy. With a nod to Mayr, in this paper, we present an operational criterion to demarcate sentient purposeful systems from mere historical end-directed processes. The main issue here is that goal-directedness could be used equivocally and applied to both historical processes, such as evolution or even some physical processes whose goal is to follow the laws of nature, such as the laws of thermodynamics, and in a more specific sense, to instances of human agency. Although the terms 'agency' and 'teleology' can be used equivocally to describe different sorts of end-directedness or goal-directedness in natural or artificial processes at various levels of complexity, the primary goal of this paper is to establish a distinction between what we term *genuine* agency and more rudimentary forms of goal-directedness. We use the term 'genuine' in this paper in a broad sense and for want of a better word, without insisting that other kinds of goal-directedness or teleonomy are counterfeit. What we are really concerned with is finding an objective (in the sense of scientifically informed) measure for setting a meaningful distinction between different kinds of goal-directedness that are at play across physical, biological, and psychological domains. The aim is to differentiate more sophisticated agency from less elaborate forms of teleology and agency that manifest in other kinds of natural or artificial processes. Thus, inspired by Mayr's attempt to clarify the difference between diverse kinds of end-directedness, we delve into *technical details* to elucidate the distinctions between the kinds of goal-directedness and purposiveness at different levels. Our technical framework will be presented in terms of the Free Energy Principle (FEP), which offers a comprehensive approach to understanding the behaviour of complex systems, particularly biotic self-organising systems, and provides a method for interpreting these systems as entities striving to minimise their variational free energy. Variational free energy acts as an upper limit on surprisal, quantifying the deviation between expected and observed consequences. By minimising this free energy, systems effectively reduce surprisal, implying a dynamic process of inference and belief updating. This is particularly significant for organisms, as it explains how they navigate their environment by resolving uncertainties and adapting to sensory inputs. Minimising variational free energy implicitly maximises model evidence (a.k.a., the marginal likelihood of observations under some model), facilitating a more accurate representation of the world. This process, known as active inference, describes how organisms actively engage with their environment to gather sensory evidence that aligns with their internal models. The crucial point is that, while some self-organising systems minimise only their (variational) free energy in the moment, others can also minimise their *expected* free energy by invoking models that allow for inferring the counterfactual outcomes of actions. We will offer more technical details later in this paper, but for the time being, suffice it to say that we ascribe a human-like sense of agency and autonomy to self-organising systems that possess the ability to form counterfactual models of the outcomes of their actions, but not to other kinds of self-organising systems. These self-organising systems with thick generative models can be understood as sophisticated conservative particles, as opposed to simple or ordinary ones. For example, equipped with a model of the consequences of action, 'goals' become intended states in the future that may or may not be the states that are realised; on this view, you and I can have 'goals', but the weather cannot.

It is also worth mentioning that while we mainly build on Mayr's distinction between different kinds of goal-directedness, our proposal also speaks to the cybernetic approaches to goal-directedness; especially as the FEP offers a modern expression of fundamental cybernetic concepts such as homeostasis and control (Buckley et al., 2017; Friston et al., 2010; Seth, 2014). For instance, our proposal aligns with Wiener's (1948) view on goal-directedness (or purposiveness) along the lines of negative feedback mechanisms.[2] However, we strive to go further and advance a demar-

[2] Negative feedback is a regulatory mechanism in self-organising systems across biological and physical domains. It operates to measure and amend the discrepancy between desired and actual inputs. When a system deviates from its optimal condition, negative feedback mechanisms act to restore equilibrium (or homeostasis if it is in the biological domain). According to Wiener (1948, Chapter IV), the negative feedback mechanisms govern the goal-directedness of natural biological systems as well as designed ones. For example, a fever is something that encroaches on the body's homeostasis, and the rise of the body temperature is generally a sign of illness, and a permanent variation of five degrees is scarcely consistent with life. So, when the fever rises, negative feedback mechanisms are triggered; the hypothalamus in the brain activates cooling responses and dilation of blood vessels near the skin surface to release heat, and there will be sweating to evaporate heat. The same negative feedback mechanisms are at work to regulate inanimate objects, such as the working of the thermostat, which turns the heater on when the room temperature falls below a degree (Wiener, 1948, Chapter IV). There are cases of goal-directedness across

cation criterion to differentiate between forms of goal-directed behaviour reliant on negative feedback—in terms of merely responding to external influences—and more sophisticated forms of agentic goal-directedness that are associated with planning and decision-making that are informed by future-oriented beliefs, shaped by internal states. In the same vein, we appreciate Ernest Nagel's view on the prevalence of goal-directedness across various domains, regardless of whether it involves human agents with intentions, living systems without intentions, or inanimate systems (Nagel, 1977). Nagel's view has also been explicated in terms of cybernetics[3]: assuming that every process in which some equilibrium state is restored would also have to be designated as goal-directed, regardless of whether it is natural or artificial (Nagel, 1977, p. 274). While this general and foundational notion of goal-directedness is quite congenial to our project in this paper, our endeavour also aims to provide an ontology of kinds of goal-directedness.[4] To improve on the abovementioned accounts, we remark that in the context of the FEP, conservative particles are identified here in contrast to dissipative particles. Unlike particles that dissipate under the force of the environment, conservative particles elude random fluctuations just because they have precise dynamics. While this property bestows some level of goal-directedness to conservative particles—in following paths of least action—it does not imply that all conservative particles have a human-like kind of agency that rests on inferring the consequences of their action. We will decipher this proposal later in this paper by referring to technical details (e.g., in Friston et al., 2022). However, before delving into such details, we provide a concise exposition of recent scientific moves that reinforce the notion of teleonomy in terms of programmability.

Echoes of Mayr's proposal—to characterise teleonomy in terms of programmability—can be found in recent endeavours to define teleonomy by leveraging the computational framework of bioelectricity. This concerns the scaling of cell computation into anatomical homeostasis and the evolutionary dynamics of multi-scale competency (Clawson & Levin, 2022; Levin, 2022b; Levin & Martyniuk, 2018). Both Mayr's and Levin's views converge in their specification of goal-directedness by emphasising the importance of information codes, computation, and programming. Mayr's perspective highlights the significance of programmability in distinguishing teleonomic systems; underscoring the role of information processing and computational capabilities in generating purposeful behaviour. In the same vein, Levin's work

designed and natural domains, insofar as they use these mechanisms to achieve homeostasis in their environment. In this context, it has been argued that the purposiveness of some artefacts, such as heat-seeking missiles, is intrinsic and does not depend on the conscious intentions of the designer (Garson, 2016, p. 19; Rosenblueth et al., 19431943).

[3] According to Nagel (1977a), in order to be recognised as goal-directed systems, entities across natural and artificial domains have to satisfy two conditions; *plasticity* and *persistence*. The first condition involves showing adaptability, while the second condition consists of systems maintaining their trajectory toward the goal over time.

[4] While Nagel generally argues that there may not be anything inherently special about goal-directedness in the biological domain, it has been contended that providing viable explanations in biology may necessitate considering the goal-directedness of biological functions, without reducing them to mere causal mechanisms at the physical level (Cartwright, 1986). Our proposal here offers a reconciliatory approach by acknowledging the prevalence of goal-directedness across diverse domains, while also clarifying the subtle differences between goal-directedness at play across them.

identifies bioelectric networks as the computational medium underlying teleonomy in biological processes, elaborating on how "evolution exploits the laws of physics and computation in the context of teleonomic processes" (Clawson & Levin, 2022, p. 16). Thus, they converge on a position that teleonomy can be understood through the lens of computational principles and the manipulation of information. Amidst the points of convergence also stands a disagreement between Mayr's perspective and Levin's: Mayr strongly asserts a distinction between teleonomic processes and those that lack teleonomy and elaborates on multiple meanings of teleonomy. Conversely, Levin adopts a more general (and also seemingly monolithic) stance towards the concept of teleology/teleonomy; arguing that it serves as a deep principle for understanding various aspects of biology, and identifies it with the "nested goal directedness at multiple levels" (Clawson & Levin, 2022, p. 4). His proposal portrays a continuum of cognitive processes—including agency and teleonomy—across multiple scales, and by the same move brushes away the boundaries between systems that exhibit true purposiveness and agency and those that lack such characteristics. For instance, consider a population of cells that collectively work towards the goal of organising themselves into tissues and organs. However, their purposive behaviour does not seem to evince the same kind of agency that can be observed in whole organisms or human-like creatures. To anticipate, we will improve on Levin's point in this paper by arguing that not all hierarchical structures embody genuine instances of goal-directedness and agency. Instead, teleology and agency emerge within multiscale systems that are equipped with temporally deep structures enabling the organism to minimise not only its variational free energy but also expected free energy. This is based on the ability to model the counterfactual outcomes of its actions The field of theoretical biology has long been concerned with making classifications, creating categories, constructing genealogical trees, and delineating the boundaries of domains of life (Williams & Ebach, 2020). In this spirit, our objective to carve teleology at its joints not only aligns with Mayr's differentiation between genuinely purposive agents and non-purposeful historical processes but also supplements Levin's work with a criterion for distinguishing between different kinds of goal-directedness across different physical, biological, and psychological scales.

3 Carving the structure of nature

The attempt to have a well-articulated measure for carving nature at its joints is well-motivated, primarily driven by scientific imperatives. Our proposal is supported by a strong scientific motive: scientists across disciplines consistently engage in objective classification tasks. In biology, taxonomic methodologies classify organisms, while chemistry relies on the periodic classification of elements. Physics categorises particles into distinct types. Additionally, philosophical motivations contribute to our proposal.

Concerning philosophical motives, we are sympathetic to the pragmatic value of viewing teleology as a lens for studying morphologies, and it may well be the case that the question of whether systems are *really* teleonomic is a red herring (Levin, 2022a). However, even from a Dennettian perspective—which Levin (2022a, b) adopts—not all scientific patterns, discerned through an intentional stance, hold the same epistemic and ontological weight.[5] It becomes a viable philosophical project to delineate instances of (relatively) genuine cases of agentic goal-directedness from those that demonstrated a lesser degree of agency. Indeed, the notions of "carving" and "joints" jump to mind. Our aim is not to delineate fixed categories of teleological phenomena, but rather to develop a scientific measure that allows us to assess and specify separable kinds of agency and teleology across various domains, both natural and artefactual.

4 The free energy principle

The account of the emergence of intelligence and agency in collective systems is grounded in the understanding that these systems assess and maintain their homeostatic activity through hierarchical (active) inference, thereby minimising variational free energy. This has been foreshadowed in cybernetics, to account for how negative feedback mechanisms contribute to homeostasis (Garson, 2016; Seth, 2014; Wiener, 1948). Minimisation of free energy takes place with various degrees of complexity at different scales or levels within systems, demonstrating a comprehensive approach that' eludes a specific hierarchical scale. Under the minimisation of free energy, higher-level goal-directed agents shape the option space available to lower-level agents. This configuration of the option space empowers lower-level agents to navigate their goals by minimising free energy, where the influence of higher-level agents is instrumental in shaping the navigational capabilities and contextualising goal attainment of lower-level 'agents' (Levin, 2022b). The Free Energy Principle (FEP) and its corollary, active Inference (Friston, 2010, 2012; Pezzulo et al., 2022), offer a formal framework for naturalising the concepts of agency and goal-directedness, in terms of the physics of self-organisation and a basic kind of sentience.

Levin's insights into the functioning of collective intelligent agents align well with the FEP, demonstrating how their behaviour can be understood within the context of minimising variational free energy and navigating goal-directed spaces. Furthermore, Mayr's classical concept of teleonomy, which emphasises programmability, finds renewed strength and support through the lens of FEP. In anticipation, our argument will posit that a system can genuinely exhibit teleology if the generative models it entails possess temporal depth. This deep structure entails the presence of Markov blankets, which define the hierarchy; enabling the modelling of counterfactual consequences of potential actions. The particular kind of depth we appeal to here is temporal depth; namely, the ability to anticipate the future and model the (plausibly

[5] See the relevant discussion in terms of the computational depth or projectability of patterns, which establishes a distinction between patterns in a general sense and what Dennett refers to as "real patterns"–albeit without insisting that real patterns are *really* real (Beni, 2017; Dennett, 1991; Ladyman & Ross, 2007). Indeed, a similar principle can be extended to the concept of teleonomy. In this context, it is worth noting that, arguably, computational depth, algorithmic complexity, universal computation, and information entropy are all manifestations of the Free Energy Principle (Friston, 2010; Friston et al., 2022). Upcoming research may focus on technical details of how this principle can be similarly extended to the concept of teleonomy/teleology.

Markovian) world as quintessentially non-Markovian. Specifically, the future of the system in question depends upon its action, which endows the system with a minimal kind of agency; namely, a system that acts upon the world in a future-pointing fashion. Importantly, non-Markovian dynamics are a necessary aspect of any hierarchal generative model; in the sense that a deep generative model entails a separation of temporal scales; e.g., (Friston et al., 2018). Such a configuration bestows upon the system the capacity to plan and strategize, thereby opening avenues for goal-oriented behaviour.

The FEP provides a method (i.e., physical principle) for understanding how biotic self-organising systems, including collective intelligent systems (and even semi-biotic ones, for that matter), can be read—or simulated—as minimising their variational free energy. The variational free energy acts as an upper bound on the surprisal (i.e., surprise or self-information), which is the negative logarithm of (Bayesian) model evidence (Friston, 2012). To provide an interpretation of surprisal in terms of model evidence, it is necessary to condition surprisal on a generative model, which is just the joint density over (unobservable) causes and their (observable) consequences (Friston et al., 2020). In other words, the generative model prescribes the probabilistic generation of consequences (e.g., sensations) from causes (e.g., latent states of the world). Inverting such models by minimising variational free energy—or maximising Bayesian model evidence (a.k.a., marginal likelihood) can thus be read as inferring (unobservable) causes from (unobservable) consequences. This process can be neatly summarised as minimising self-information or maximising model evidence; namely, self evidencing (Hohwy, 2016).

Surprisal or self-information is a measure of the deviation or discrepancy between the expected consequences predicted by the model and the observed consequences (in some settings, this can be described in terms of prediction errors). Information entropy is the time average of surprisal. By minimising their variational free energy, the systems implicitly minimise their entropy or surprisal. Formally, for any given system or model m:

$$\mathrm{F} = -\log \mathrm{P}\left(\mathrm{o}|\,\mathrm{m}\right) + \mathrm{KL}[\mathrm{Q}\left(\mathrm{s}\right) \,||\, \mathrm{P}\left(\mathrm{s}|\,\mathrm{o},\mathrm{m}\right)] \quad (1)$$

The free energy (F) is an information-theoretic measure that binds surprise. It combines surprisal with the divergence between the probability distribution $\mathrm{P}\left(\mathrm{s}|\,\mathrm{o},\mathrm{m}\right)$ over latent states, s generating observed data, o and the approximation provided by a variational density $\mathrm{Q}\left(\mathrm{s}\right)$ parameterised by internal states. The term $-\log \mathrm{P}\left(\mathrm{o}|\,\mathrm{m}\right)$ is the surprisal associated with the observed data under the generative model. Minimising the free energy implicitly minimises the KL divergence and therefore improves the accuracy of the variational approximation. This is why minimising variational free energy can be read as inference. If the system now acts upon the world to change observations or outcomes, minimising free energy corresponds to minimising surprisal. This is active inference.

Organisms can therefore be described as minimising surprisal via a process of (Bayesian belief) updating of their internal states to better explain observed sensory inputs. By doing so, they effectively resolve uncertainty about unobservable states of their external milieu; providing a principled explanation for perceptual inference in things like the brain. Moreover, the minimisation of variational free energy, which maximises Bayesian model evidence $\mathrm{P}(\mathrm{o}|\mathrm{m})$, not only accounts for perceptual inference but also explains adaptive action and behaviour. This speaks to the notion of active inference; namely, the notion that organisms actively engage with their environment to acquire sensory evidence for their internal models or beliefs that minimises the discrepancy between their models and the environment by *changing the environment*.

The FEP foregrounds the intimate relationship between perception and action, wherein organisms actively gather information from their surroundings to minimise surprisal. In cases where there is a mismatch between sensory evidence and internal models, organisms can update their internal states to resolve the mismatch or alternatively modify their environment to better match their internal models (Sajid et al., 2021). For example, when an organism senses excessive warmth—which is a mismatch between the predicted and observed sensory inputs related to temperature—by updating its internal models, the organism can bring its Bayesian beliefs in line with the observed sensory inputs (i.e., it can infer that it is hot). Alternatively, the organism may seek out shade to cool itself (i.e., so that its body temperature returns to that which was predicted). In short, through an interplay of perception and action, FEP explains the capacity of complex systems, including the brain, to counteract fluctuations in a capricious and itinerant world—and uphold a state of sustained and homeostatic interaction with that world (Pezzulo et al., 2015). In short, by minimising variational free energy, these systems are able to effectively navigate their surroundings and maintain stability by resolving surprise and uncertainty.

As we remarked before, homeostatic interaction with the world—in terms of variational free energy—aligns with classical cybernetic accounts, where homeostasis in biological systems is achieved through negative feedback (where deviation from the desired state (in designed as well as natural systems) is remediated through negative feedback mechanisms (Wiener, 1948) (also see footnote 2 in this paper). Accordingly, the FEP-based account of goal-directedness aligns with Wiener's theory of purposefulness. Against this backdrop, we offer a formal measure to demarcate different kinds of goal-directedness. This refinement is based on the general insight that certain complex systems can be deemed more agentic based on the depth or thickness of their generative models[6]. This temporal depth allows organisms to go beyond immediate sensory inputs and incorporate counterfactual predictions, enabling them to extend their predictions and anticipate future outcomes based on distinct courses of action (Parr & Friston, 2019). By considering the depth of generative models, we therefore naturalise the ability of certain kinds of organisms to evince anticipatory behaviour; thereby, enhancing their adaptability and decision-making processes. By

[6] Technically speaking, certain systems come to represent their own actions as causes of their sensory observations. For these kinds of systems, their most likely dynamics minimise a free energy functional called expected free energy. It is called expected because it is the expectation under the most likely outcomes in the future) (Friston et al., 2022).This means that the behaviour of such systems can be cast in terms of planning as inference: (Attias, 2003; Botvinick & Toussaint, 2012; Parr & Friston, 2018a). In short, particular kinds of systems are distinguished from other kinds by their capacity to plan; endowing them with minimal form of agency and apparent purpose—quantified by the characteristic states that underwrite the system's generative model.

considering multiple possible trajectories and actions, sophisticated systems of the sort (look as if they) evaluate the consequences of different choices and select optimal courses of action (Parr & Friston, 2018a, b). In summary, the depth of generative models endows behaviour with purposeful aspects that license the notion of *agency* and move beyond self-organisation based purely upon historical (Markovian) processes. One might argue that temporal depth is necessary for ascribing to them the attributes of intentionality, agency, and sentience (Beni, 2023; Clark, 2020; Friston, 2018).

5 A markovian articulation

As we have seen thus far, Levin's research supports the notion that teleology/teleonomy is widespread in collective intelligent systems. Levin further proposes that the teleonomy of (collective systems) involves the minimisation of (variational) free energy through active inference, enabling systems to maintain homeostasis. We quickly reviewed FEP in the previous section. Now, we turn our attention to laying the groundwork for elucidating how the FEP framework can offer a measure for quantifying the kinds of goal-directedness. To embark upon this enterprise, we will delve into some technical aspects of Markovian models within the context of FEP.

The construct of Markov blankets has been most clearly established in Bayesian network models and graph theory, as pioneered by Pearl (1988, 2000). A Markov blanket refers to a specific set of variables that shields a random variable from the direct influence of all other variables in a probabilistic graphical model. For a node within the graphical model, its Markov blanket comprises its parents, children, and co-parents of its children. This set of variables collectively forms the Markov blanket as they underwrite the conditional independencies that shield what is inside from the influence of other variables on the outside.[7]

The FEP suggests that organisms minimise their surprise or uncertainty by updating their generative models or by acting upon—and thereby modifying—their environment. This rests upon a statistical separation between internal and external states. The requisite partitioning and sparsification can be stated in terms of Markov blankets. Markov blankets can be further partitioned into sensory states that influence—but are not influenced by—internal states, and active states that influence—but are not influenced by—external states (Hipólito et al., 2021).[8] A self-organising system

[7] It is important to note that while Markov blankets provide a useful modelling framework, caution must be exercised to avoid interpreting the features of the Markovian models as literal representations of the target systems (Beni, 2021a). To be clear, indeed, the scientific credibility of Markov models of complex systems cannot be completely disregarded solely because they do not provide a literal representation of all aspects of their target systems (Kirchhoff et al., 2022). However, it is important to recognize that these models are inherently idealised and contain distortions that prevent us from drawing direct epistemic or ontic conclusions about the nature and scope of agency and teleonomy in complex systems.

[8] This definition presumes the principle of least action, which is a variational principle that states that the path taken by a system between two points in space and time is the one for which the action is stationary (i.e., maximal or minimal). In this context, action is defined as the integral of the Lagrangian over time, from the initial point to the final point. The FEP characterises the optimal trajectory that minimises the (path integral of) free energy functional for a given initial state and a generative model. In short, the FEP

(e.g., a particle, person or population) is defined as comprising internal and blanket states, where the blanket states act as a boundary between the internal and external states. This structure implies that the internal paths of the particle are conditionally independent of the external paths, given the blanket paths. This conditional independence implies the existence of a most likely internal path and a posterior probability density over external paths, *for each blanket path*. In turn, this means there is a mapping between internal dynamics and a probabilistic representation (a.k.a., Bayesian belief) over external states. It is this mapping that licences an interpretation of self-organisation in terms of inference and self-evidencing (Hohwy, 2022; Ramstead et al., 2022).

The ensuing inferential process underwrites Bayesian beliefs about the external paths. The paths of least action for conservative particles, which are sufficiently large, correspond to the most likely internal paths and minimise (the path integral of) self-information or surprisal (Friston et al., 2022, p. 5). In the context of the present discussion about teleonomy/teleology, it should be noted that in the case of advanced teleological agents, the path integral can be decomposed into the expected surprisal of sensory paths minus expected information gain. This decomposition provides an interesting interpretation where such paths appear to minimise expected surprisal, while maximising expected information gain; i.e., minimising uncertainty and surprises. More on this in the next section.

It is worth mentioning that Markov blankets do not directly map to specific neurobiological cell boundaries. Instead, they facilitate dynamic communication across scales (Hesp et al., 2019; Hipólito et al., 2021) or rather represent the patterns of interactions between internal and external states, such as ion channel conductance, through variables like membrane potentials. The point here is that the boundaries of a Markov blanket do not (necessarily) correspond to the boundaries of traditional biological entities but offer a means to consider the boundaries of biological systems—ranging from individual cells to individual organisms, and societies of natural and artefactual organisms—in terms of their inputs and outputs (i.e., sensory and active blanket states).

This take on minimising free energy is compatible with Levin's conception of teleonomic systems, as it holds that the boundaries of such systems need not align strictly with the biophysical boundaries of a living organism. In other words, teleonomic systems can be hierarchically composed of Markov blankets of Markov blankets, and there is agency and teleonomy all the way down (Palacios et al., 2020). This hierarchy extends downward to individual cells and extends outward and upward to include not only individuals but also large collective systems (Badcock et al., 2019; Hesp et al., 2019; Ramstead et al., 2018), such as societies and populations of individuals (Kirchhoff et al., 2018). This speaks directly to Levin's account of how teleology/teleonomy across multiple scales relies on the combined functionality of ion channels

just is a principle of least action, which states that biological organisms and cognitive systems can always be described as adjusting their internal states and behaviours to match their expectations with sensory input, thereby maintaining a nonequilibrium steady-state while in exchange with their environment (via their Markov blanket).

and pumps within collective systems (Levin et al., 2019).[9] One can plausibly assume that Markov blankets provide a tool for representing the multi-scale dynamics of systems as envisioned by Levin, facilitating the understanding of goal-directedness and teleonomy. However, it remains to be established whether Markov blankets capture the fundamental structure of teleonomy at all levels. And whether each unit and scale within a nested system are teleonomic on an equal footing. We shall unpack this point immediately.

Indeed, the concept of hierarchical depth entails the existence of nested levels within a system, where higher-level representations and processes exert top-down influences on lower-level components (Badcock et al., 2019; 2018; Hesp et al., 2019). In this context, Markov blankets are organised hierarchically, with each Markov blanket existing within a broader or encompassing Markov blanket (Fields et al., 2022). Nested Markov blankets indicate that variables at higher levels of the hierarchy have a broader scope of influence and control, encompassing and regulating the variables at lower levels. Crucially, the course-graining that necessarily accompanies movement from one scale to the higher scale induces the separation of temporal scales such that things at higher (coarse-grained) scales change more slowly. This hierarchical organisation is often observed in neural systems and cognitive architectures, enabling the integration of information and the generation of scale-invariant (and sometimes scale-free) complex behaviours.

However, it is essential to note that not all hierarchical structures possess advanced teleological properties, and they do not necessarily exhibit agency, unless there is temporal depth at each (or at least some) level of such systems (how to individuate such multiscale systems is yet another important question). In other words, the presence of temporal depth is crucial for authentic teleology and agency. For instance, collective systems—like natural selection at a superordinate scale or a single cell at the subordinate level—may embody some hierarchical depth but lack advanced agency or teleology. On the other hand, there are advanced forms of teleology and agency found at the level of phenotypes. In such systems their nested hierarchical structure embeds Markov blankets with temporal depth, allowing them to exhibit authentic teleological properties and agency. By leveraging Markovian models, capable of representing counterfactual outcomes of future action sequences, such systems can simulate and assess potential future scenarios, with a separation of temporal scales. The separation of timescales engenders planning and the formulation of long-term (non-Markovian) strategies. In short, the ability to plan and strategize is directly related to the counterfactual depth of models that represent the future consequences of actions. In this context, we submit that advanced teleologic systems are those taking this act further

[9] To clarify this point: arguably communication and functionality of ion channels can be effectively modelled using the Markov blanket formulation, particularly in terms of a Markovian formulation of the Hodgkin-Huxley model of bioelectric communication, where informational communication can be represented by capturing the influences of external states (e.g., electrophysiological setup) via sensory states (injected current) on internal states (ion channels). These internal states, in turn, influence the active states represented by the membrane potential. The application of Markov blankets enables a comprehensive understanding of the dynamics of communication across different scales, where the capacity of the channel in the context would mediate the influence of the injected current and ion channel currents on the membrane potential (Hipólito et al., 2021, pp. 90–91).

by incorporating self-modelling, enabling them to form estimations of which possible course of action in the future would most effectively resolve their environmental uncertainty. We will flesh out this insight further in the next section.

6 Multiple meanings of goal-directedness

In this section, we provide a measure for the diverse kinds of goal-directedness, grounded in the technicality of FEP. To that effect, we offer a couple of demonstrations that operationalise some of Mayr's insights.

In his work on the Multiple Meanings of 'Teleology', Mayr (1998) distinguishes five ways of speaking of teleonomy. These distinctions are more refined than the earlier (1961) differentiation between historical processes and teleonomic systems. According to Mayr, multiple senses of teleology apply to (a) processes in inanimate nature whose end-directedness is determined by universal laws of nature, such as the law of universal gravity and the second law of thermodynamics, (b) teleonomic processes driven by programs, (c) adapted features resulting from natural selection rather than executive processes, (d) purposeful behaviour, exemplified by the deliberate actions of a pride of lionesses that strategically divide into two groups to attack prey—e.g., one group launches a direct attack, while the other cuts off the prey's escape route, and (e) cosmic teleology, the end-directedness of nature tending towards perfection. We demonstrate our point by setting a technical distinction between some of the kinds of goal-directedness that are mentioned by Mayr.

Let us start with what Mayr calls teleomatic systems, e.g., a cup of tea losing its heat in a cold environment. This can be represented by the principle of least action, which suggests that a system tends to follow a path that minimises its overall action. A cup of tea, for example, loses heat over time to reduce the temperature difference between itself and the colder environment. To relate this idea to the paper's topic, we can relate the phenomenon to the concept of variational free energy. However, to model this process, we do not need to attribute expectations about the consequences of actions to the cup. The process can simply be modelled based on implementing the laws of thermodynamics. This, however, does not imply that the process lacks a form of goal-directedness. Assuming the goal of the cup is to achieve thermal equilibrium with its environment, negative feedback mechanisms can be utilised to model how it gradually approaches the ambient temperature. By adhering to thermodynamic principles and employing negative feedback loops, the tea adjusts its temperature accordingly. From Wiener's (1948) point of view, there is some form of purposiveness in the cup of tea's process of achieving thermal equilibrium with the environment. And the FEP can describe that level of goal-directedness in terms of negative feedback mechanisms. However, the FEP provides a framework for disambiguating subtle differences between various kinds of goal-directedness. For instance, the process involved when a cup reaches thermal equilibrium with the environment differs from more complex (non-Markovian) forms of goal-directed behaviour, such as when a sentient being attempts to administer antipyretic medications to reduce a fever. The main point is that the cup doesn't need to model itself or plan its actions to reduce the difference between its internal and external temperatures—the equilibration is just a manifestation of the cup's existence within a particular context, where the joint cup-context system has some characteristics states (i.e., nonequilibrium steady-state). The process is simply a result of teleomatic interactions that are not based on internal influences on the cup's part. To provide a reliable representation of this situation, there is no need to model the cup as something capable of forming internal states, such as expectations about the consequences of its actions; instead, it merely responds to external influences (Friston et al., 2022).

On the other end of the spectrum, the same equation could be used to model a cosmic process, or, rather a historical one, such as evolution, where organisms adapt their genetic and phenotypic states to match selective pressures, reducing prediction errors and increasing fitness (Campbell, 2016; Frank, 2012; Friston et al., 2023; Vanchurin et al., 2022). To support the claim that systems representable by Markovian models exhibit teleology and agency, it is sufficient to consider that natural selection and genetic drift are guided by implicit foresight towards potential outcomes that maximise species' survival in a cost-efficient manner. It is possible to imagine that natural selection operates based on expectations regarding the role of an organism's genotype in carrying useful genetic information to the next generations. In cases where a specific allele does not align with anticipated future outcomes, natural selection addresses the discrepancy by investing in alternative genetic variations (Kauffman, 1993; Voigt et al., 2000; Woods et al., 2011). This adaptive strategy reduces uncertainty and increases confidence in the fitness or survival rate associated with specific traits in descendant populations (Beni, 2021b). In the same vein, from the assumption that evolution exhibits an exceptional capacity for rapid and effective utilisation of mutations to exploit structural and functional opportunities, it may follow that evolution employs a prediction error-minimising architecture[10] that eliminates the need for excessive training on past conditions, where evolving becomes a matter of using intelligent collective mechanisms that inherently demonstrate resilience, adaptability, and triumphant adaptation in entirely novel circumstances. Pursuing this line of reasoning, Levin (2023, p. 6) argues that "the space which evolution actually searches is not only the space of microstates of the genome but also a much more tractable space of behaviour shaping signals: evolution exploits cellular intelligence as a highly exploitable affordance". However, ascribing foresight to natural selection implies a level of intentionality that may not apply to evolutionary processes (from our perspective, because such historical processes lack temporal depth and non-Markovian properties). In short, such Markovian processes—such as a cooling cup of tea or evolution—exhibit end-directedness without having insights into the counterfactual consequences of their actions. They lack temporal depth. On the other hand, systems such as human beings, and other animals that can display anticipatory behaviour, do possess advanced forms of teleology and agency. These organisms possess the ability to foresee future consequences of their actions, embedding them via the temporal depth of their models. The temporal depth in such systems is manifested in the property of conservatism, allowing them to counter random fluctua-

[10] Noting that prediction errors are, technically, free energy gradients. This means that any free energy minimising process—such as natural selection or Bayesian model selection—based on variational free energy or marginal likelihood (a.k.a., model evidence) can be read as minimising prediction errors.

tions.[11] Physiologically speaking, this looks very much like a move from homeostasis to allostasis (Corcoran et al., 2020; Seth and Friston, 2016; Stephan et al., 2016; Sterling & Eyer, 1988); namely, pre-emptive, (agentic) behaviour that eschews the need for (Markovian) homeostatic responses. Additionally, being sufficiently large and sparsely coupled, these systems effectively segregate active states from internal states, enabling their capacity for being agentic (Friston et al., 2022) and as we submit here, also agentic teleology.

In other words, somewhere in between cups of tea and evolution, there exist agents that are capable of exhibiting advanced teleology, where their goal-directedness can be explicated in terms of planning as inference. In the case of these 'strange' particles[12]—to borrow a term from (Friston et al., 2022)—intentional and purposeful forms of goal-directed behaviour are associated with insights into counterfactual outcomes of their actions. A subtle distinction between conservative (Markovian) particles—that follow precise, paths of least action—and strange (agentic) particles is that the actions of agents are incorporated into their generative models (i.e., models that can generate predictions). In other words, action is treated as a cause of sensory consequences. This follows naturally when—for certain kinds of particles—active states do not influence internal states directly. The most likely paths (of least action) for conservative (Markovian) particles are those that minimise expected free energy. For strange (agentive) particles, this minimisation becomes an explicit part of the generative model: i.e., the agent 'believes' her unobservable actions will minimise expected free energy, while action per se realises those beliefs by minimising prediction errors or variational free energy.

The ability to minimise expected free energy—characterised as a combination of risk and ambiguity (Parr & Friston, 2019)—plays a pivotal role in this process. Humans, for instance, depend on the depth of their generative models to choose strategies that lead to optimal outcomes and effectively resolve uncertainties based on counterfactual possibilities. Expected free energy can be further comprehended as the subtraction of expected information gain from expected cost. By contemplating alternative actions and their potential consequences, individuals can quantify the reduction in surprise or uncertainty in their surroundings and make informed decisions. This interplay between expected cost and expected information gain serves as a guiding principle for optimising behaviour and aligning actions with desired outcomes (Friston et al., 2022).

Consider the case of a person who strategically adds cool milk to her tea. By leveraging her understanding of the counterfactual outcomes of her actions, she aims

[11] Conservative particles (in contrast to dissipative particles) are active (in contrast to inert) particles that follow the path of least action. Technically, this means that, a conservative particle "is an active particle whose random fluctuations on particular states have amplitudes that are infinitesimally small" (Friston et al., 2022, p.16).

[12] This means that, "the autonomous paths of these kinds of particles have a Lagrangian, known as expected free energy, which can be decomposed into terms corresponding to expected cost and expected information gain. In this setting, cost is the Lagrangian or surprisal of sensory paths, which defines the characteristic trajectories of a particle." (Friston et al., 2022, p. 3). Such a strange particle not only minimizes its free energy but also is capable of inferring its own actions and believing that it is a conservative particle (ibid.).

to achieve the desired result. In this scenario, the person's expected cost is the surprising sensation of drinking tea that is too hot. On the other hand, the expected information gain of gently sipping her tea to test its temperature may supervene, if she believes the tea will not scald her. By subtracting the expected information gain from the expected cost, the person can assess the trade-off between the potential benefit (reducing uncertainty) and the potential cost (a scalded tongue). As the example indicates, in the case of such sophisticated organisms, it is important to note that their internal states are not solely influenced by their actual actions alone. They are also influenced by their beliefs about the relationship between their possible actions and the environment, as well as the anticipated consequences of those actions: e.g., the tea drinker's internal states are not only influenced by the act of adding cold milk but also by their beliefs about how this action will affect the temperature of the tea and whether she can drink it. In short, expected free energy brings affordances (Bruineberg et al., 2018; Gibson, 1977) to the table in a way that could be considered definitive of agency and purpose. Expectations about the consequences of actions play a significant role in shaping internal states and subsequent behaviour. This point has been foreshadowed in the free energy account of self-consciousness:

> We elude the problems of calling evolution conscious, because the process of natural selection minimises surprisal (i.e., maximises adaptive fitness) but not expected surprisal or uncertainty (i.e., adaptive fitness expected under alternative evolutionary operations or selection). The key difference between (self) consciousness and more universal processes then appears to be the locus of selection. In non-conscious processes this selection is realised in the here and now with selection among competing systems (e.g., phenotypes). (Friston, 2018, p. 5)

This approach avoids the metaphysical conundrum of backward causation[13] because acknowledging that a sophisticated cognitive organic system, such as a sentient being, is capable of anticipating the counterfactual consequences of her action does not rely on reversing the temporal order of cause and effect or resorting to supernatural or mysterious elements. Instead, it arises from the implicit coupling between beliefs about the future and current actions. In this sense, the future can be said to "cause" the past, although not in the conventional sense of causing events to occur before their causes (Parr & Friston, 2019). Instead, agents rely on internal states that parameterise beliefs about potential future outcomes. These internal states—encompassing the agent's prior experiences, current observations, as well as expectations about the consequences of their actions—influence policy selection through the minimisation of expected free energy. This minimisation process drives the agent towards behaviours that align with its internal model and reduce the discrepancy with the observed reality (Parr & Friston, 2019). In doing so, the agent's behaviour becomes non-Markovian, as it takes into account counterfactual outcomes of its own future courses of action. The primary focus here is on how the agent's deep generative models encapsulate its expectations, regarding the consequences of future actions. The explanation

[13] See (Faye, 2015) for a concise account of backward causation.

of backward causation that arises does not imply that the future genuinely influences the past. On the other hand, the teleomatic goal-directedness of any other historical process can be modelled as Markovian (as opposed to non-Markovian) because they do not possess internal states that influence their future behaviour. For instance, the rate at which the tea cools depends solely on its current temperature difference compared to the surrounding environment, and this dependence on the current state directly determines the future state (e.g., cooler temperature). Even though such processes remain goal-directed in a basic sense, e.g., in the sense of showing some sort of persistence to follow a specific trajectory to reach thermal equilibrium with their environment, they lack the ability to model the consequences of their actions in the future or form expectations about such consequences and use that as the basis of their goal-directedness. Thus, we suggest demarcating between austere forms of goal-directedness and the intricate and purposeful teleonomic behaviours of sentient beings.

7 Concluding remarks

In this article, we drew on the Free Energy Principle (FEP) to provide a scientifically cohesive measure to distinguish different kinds of goal-directedness in self-organising complex systems, with an eye to Ernst Mayr's classical conceptualisation of the same idea. By identifying agential teleology with certain (strange, agentic) systems that are capable of engaging in planning as inference, and contrasting them with historical (conservative, Markovian) self-organising processes, we can characterise teleologic phenomena with generative models that evince temporal depth. This enables us to differentiate anticipatory behaviour in certain organisms from the historical processes observed in natural phenomena like evolution, all contributing to a comprehensive understanding of goal-directedness across various complexity scales.

Acknowledgments We thank the reviewers of this journal for their useful and constructive comments. We would also like to thank members of the UCL Theoretical Neurobiology Group, particularly Maxwell Ramstead, for their feedback and suggestions on the presentation of this work.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Attias, H. (2003). Planning by probabilistic inference. *International Workshop on Artificial Intelligence and Statistics*, 9–16.

Auletta, G. (2011). Teleonomy: The feedback circuit involving information and thermodynamic processes. *J Mod Phys*, *2*, 136–145.

Beni, M. D. (2017). Structural realism, metaphysical Unification, and the ontology and epistemology of patterns. *International Studies in the Philosophy of Science*, *31*(3), 285–300. https://doi.org/10.1080/02698595.2018.1463691.

Beni, M. D. (2021a). A critical analysis of Markovian monism. *Synthese 2021 199:3*, *199*(3), 6407–6427. https://doi.org/10.1007/S11229-021-03075-X.

Beni, M. D. (2021b). A free energy reconstruction of arguments for panpsychism. *Phenomenology and the Cognitive Sciences*, *22*(2), 399–416. https://doi.org/10.1007/S11097-021-09739-W/METRICS.

Beni, M. D. (2023). The intentional structure of generative models. *Phenomenology and the Cognitive Sciences*, 1–12. https://doi.org/10.1007/s11097-023-09914-1.

Blackiston, D., Kriegman, S., Bongard, J., & Levin, M. (2022). *Biological Robots: Perspectives on an Emerging Interdisciplinary Field*. https://arxiv.org/abs/2207.00880v1.

Blackiston, D., Lederer, E., Kriegman, S., Garnier, S., Bongard, J., & Levin, M. (2021). A cellular platform for the development of synthetic living machines. *Science Robotics*, *6*(52). https://doi.org/10.1126/SCIROBOTICS.ABF1571/SUPPL_FILE/ABF1571_MOVIE_S2.MP4.

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, *16*(10), 485–488. https://doi.org/10.1016/j.tics.2012.08.006.

Bruineberg, J., Kiverstein, J., Rietveld, E. (2018). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese, 195*, 2417–2444.

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.

Campbell, J. O. (2016). Universal Darwinism as a process of Bayesian inference. *Frontiers in Systems Neuroscience, 10*, 49.

Cartwright, N. (1986). Two kinds of teleological explanation. *Human Nature and Natural Knowledge: Essays Presented to Marjorie Grene on the Occasion of Her Seventy-Fifth Birthday*, 201–210.

Chouraqui, F. (2011). Temporal thickness in Merleau-Ponty's notes of may 1959. *Chiasmi International, 13*, 407–427.

Clark, A. (2020). Beyond Desire? Agency, Choice, and the predictive mind. *Australasian Journal of Philosophy*, *98*(1), 1–15. https://doi.org/10.1080/00048402.2019.1602661.

Clawson, W. P., & Levin, M. (2022). Endless forms most beautiful 2.0: Teleonomy and the bioengineering of chimaeric and synthetic organisms. *Biological Journal of the Linnean Society*. https://doi.org/10.1093/BIOLINNEAN/BLAC073.

Corcoran, A. W., Pezzulo, G., Hohwy, J. (2020). From allostatic agents to counterfactual cognisers: Active inference, biological regulation, and the origins of cognition. *Biology & Philosophy, 35*, 32.

Davies, J., & Levin, M. (2023). Synthetic morphology with agential materials. *Nature Reviews Bioengineering 2023 1:1*, *1*(1), 46–59. https://doi.org/10.1038/s44222-022-00001-9.

Dennett, D. C. (1991). Real Patterns. *The Journal of Philosophy*, *88*(1), 27–51. http://links.jstor.org/sici?sici=0022-362X%28199101%2988%3A1%3C27%3ARP%3E2.0.CO%3B2-2.

Dresow, M., & Love, A. C. (2023). Teleonomy: Revisiting a proposed conceptual replacement for Teleology. *Biological Theory*, *1*, 1–13. https://doi.org/10.1007/S13752-022-00424-Y/METRICS.

Ellis, G. F. R., Noble, D., O'Connor, T. (2011). Top-down causation: An integrating theme within and across the sciences? *Interface Focus, 2*, 1–3.

Faye, J. (2015). Backward Causation. In *The Stanford Encyclopedia of Philosophy*. Stanford University. https://plato.stanford.edu/entries/causation-backwards/.

Fields, C., Friston, K., Glazebrook, J. F., & Levin, M. (2022). A free energy principle for generic quantum systems. *Progress in Biophysics and Molecular Biology, 173*, 36–59. https://doi.org/10.1016/J.PBIOMOLBIO.2022.05.006.

Fields, C., Friston, K., Glazebrook, J. F., Levin, M., & Marcianò, A. (2022). The free energy principle induces neuromorphic development. *Neuromorphic Computing and Engineering*, *2*(4), 042002. https://doi.org/10.1088/2634-4386/ACA7DE.

Fields, C., Glazebrook, J. F., & Levin, M. (2021). Minimal physicalism as a scale-free substrate for cognition and consciousness. *Neuroscience of Consciousness*, *2021*(2), niab013.

Frank, S. A. (2012). Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory. *Journal of Evolutionary Biology, 25*, 2377–2396.

Friston, K., Friedman, D. A., Constant, A., Knight, V. B., Fields, C., Parr, T., Campbell, J. O. (2023). A variational synthesis of evolutionary and developmental dynamics. *Entropy, 25*, 964.

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 (Basel Switzerland)*, *14*(11), 2100–2121. https://doi.org/10.3390/e14112100.

Friston, K. J. (2018). Am I self-conscious? *Frontiers in Psychology*, *9*, 579. https://doi.org/10.3389/FPSYG.2018.00579.

Friston, K. J., Costa, D., L., & Parr, T. (2020). *Some interesting observations on the free energy principle*.

Friston, K. J., Costa, D., Sakthivadivel, L., Heins, D. A. R., Pavliotis, C., Ramstead, G. A., M., & Parr, T. (2022). *Path integrals, particular kinds, and strange things*. https://arxiv.org/abs/2210.12761v3.

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., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews*, *90*, 486–501. https://doi.org/10.1016/J.NEUBIOREV.2018.04.004.

Garson, J. (2016). *A critical overview of biological functions*. Springer.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67–82). Erlbaum.

Gontier, N. (2022). Teleonomy as a problem of self-causation. *Biological Journal of the Linnean Society*, 1–27. https://doi.org/10.1093/BIOLINNEAN/BLAC111.

Hesp, C., Ramstead, M., Constant, A., Badcock, P., Kirchhoff, M., & Friston, K. (2019). A multi-scale view of the emergent complexity of life: A free-energy proposal. *Springer Proceedings in Complexity*, *195–227*. https://doi.org/10.1007/978-3-030-00075-2_7.

Hipólito, I., Ramstead, M. J. D., Convertino, L., Bhat, A., Friston, K., & Parr, T. (2021). Markov blankets in the brain. *Neuroscience & Biobehavioral Reviews*, *125*, 88–97. https://doi.org/10.1016/j.neubiorev.2021.02.003.

Hohwy, J. (2022). Conscious self-evidencing. *Review of Philosophy and Psychology*, *13*(4), 809–828. https://doi.org/10.1007/S13164-021-00578-X/METRICS.

Huxley, J. (1960). The openbill's open bill: A teleonomic enquiry. *Zool Jahrb (Anat)*, *88*, 1–17.

Kauffman, S. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford University Press.

Kirchhoff, M. D., Kiverstein, J., & Robertson, I. (2022). The Literalist Fallacy and the Free Energy Principle: Model-Building, Scientific Realism, and Instrumentalism. *The British Journal for the Philosophy of Science*. https://doi.org/10.1086/720861.

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.

Kuchling, F., Friston, K., Georgiev, G., & Levin, M. (2020). Morphogenesis as bayesian inference: A variational approach to pattern formation and control in complex biological systems. *Physics of Life Reviews*, *33*, 88–108. https://doi.org/10.1016/J.PLREV.2019.06.001.

Ladyman, J., & Ross, D. (2007). *Every thing must go*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199276196.001.0001.

Levin, M. (2019). The computational boundary of a "self": Developmental bioelectricity drives multicellularity and scale-free cognition. *Frontiers in Psychology, 10*, 2688. https://doi.org/10.3389/FPSYG.2019.02688/BIBTEX.

Levin, M. (2022a). *Collective Intelligence of Morphogenesis as a Teleonomic Process*. https://doi.org/10.31234/OSF.IO/HQC9B.

Levin, M. (2022b). Technological Approach to mind everywhere: An experimentally-grounded Framework for understanding diverse bodies and minds. *Frontiers in Systems Neuroscience*, *16*, 17. https://doi.org/10.3389/FNSYS.2022.768201/BIBTEX.

Levin, M. (2023). Darwin's agential materials: Evolutionary implications of multiscale competency in developmental biology. *Cellular and Molecular Life Sciences 2023*, *80:6*(6), 1–33. https://doi.org/10.1007/S00018-023-04790-Z. *80*.

Levin, M., Bongard, J., & Lunshof, J. E. (2020). Applications and ethics of computer-designed organisms. *Nature Reviews Molecular Cell Biology 2020 21:11*, *21*(11), 655–656. https://doi.org/10.1038/s41580-020-00284-z.

Levin, M., & Martyniuk, C. J. (2018). The bioelectric code: An ancient computational medium for dynamic control of growth and form. *Biosystems*, *164*, 76–93. https://doi.org/10.1016/j.biosystems.2017.08.009.

Levin, M., Selberg, J., & Rolandi, M. (2019). Endogenous bioelectrics in Development, Cancer, and regeneration: Drugs and Bioelectronic devices as Electroceuticals for Regenerative Medicine. *IScience*, *22*, 519–533. https://doi.org/10.1016/J.ISCI.2019.11.023.

Mayr, E. (1961). Cause and effect in biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist. *Science*, *134*(3489), 1501–1506.

Mayr, E. (1985). Teleological and teleonomic, a new analysis. *A Portrait of Twenty-Five Years: Boston Colloquium for the Philosophy of Science 1960–1985*, 133–159.

Mayr, E. (1988). *Toward a new philosophy of biology: Observations of an evolutionist (issue 211)*. Harvard University Press.

Mayr, E. (1998). The Multiple Meanings of Teleological. *History and Philosophy of The Life Sciences*.

Nagel, E. (1977). Goal-directed processes in biology. *The Journal of Philosophy*, *74*(5), 261–279.

Palacios, E. R., Razi, A., Parr, T., Kirchhoff, M., & Friston, K. (2020). On Markov blankets and hierarchical self-organisation. *Journal of Theoretical Biology*, *486*. https://doi.org/10.1016/j.jtbi.2019.110089.

Parr, T., & Friston, K. J. (2018a). Active inference and the anatomy of oculomotion. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2018.01.041.

Parr, T., & Friston, K. J. (2018b). The anatomy of inference: Generative models and brain structure. *Frontiers in Computational Neuroscience*, *12*, 90. https://doi.org/10.3389/FNCOM.2018.00090/BIBTEX.

Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological Cybernetics*, *113*(5–6), 495–513. https://doi.org/10.1007/S00422-019-00805-W/FIGURES/5.

Pearl, J. (1988). Probabilistic reasoning in Intelligent systems. *Probabilistic reasoning in Intelligent systems*. Elsevier. https://doi.org/10.1016/c2009-0-27609-4.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press. http://www.cambridge.org/ir/academic/subjects/philosophy/philosophy-science/causality-models-reasoning-and-inference?format=HB&isbn=9780521773621#pf97mfSQZ7HklwRV.97.

Pezzulo, G., Parr, T., & Friston, K. (2022). The evolution of brain architectures for predictive coding and active inference. *Philosophical Transactions of the Royal Society B*, *377*(1844), 20200531. https://doi.org/10.1098/RSTB.2020.0531.

Pezzulo, G., Rigoli, F., & Friston, K. J. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17–35. https://doi.org/10.1016/j.pneurobio.2015.09.001.

Pittendrigh, C. S. (1958). Adaptation, natural selection, and behavior. *Behavior and Evolution*, *390*, 416.

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger's question: A free-energy formulation. *Physics of Life Reviews, 24*, 1–16.

Ramstead, M. J. D., Sakthivadivel, D. A. R., Heins, C., Koudahl, M., Millidge, B., Da Costa, L., Klein, B., & Friston, K. J. (2022). *On Bayesian Mechanics: A Physics of and by Beliefs*. https://doi.org/10.48550/arxiv.2205.11543.

Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose and teleology. *Philosophy of Science*, *10*(1), 18–24.

Sajid, N., Ball, P. J., Parr, T., & Friston, K. J. (2021). Active inference: Demystified and compared. *Neural Computation*, *33*(3), 674–712. https://doi.org/10.1162/NECO_A_01357.

Seth, A. K. (2014). The cybernetic bayesian brain. *Open mind. Open MIND*. MIND Group.

Seth, A. K., Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 371*, 20160007.

Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A., Paliwal, S., Gard, T., Tittgemeyer, M., Fleming, S. M., Haker, H., Seth, A. K., Petzschner, F. H. (2016). Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience, 10*, 550.

Sterling, P., Eyer, J. (1988). Allostasis: A new paradigm to explain arousal pathology. In *Handbook of life stress, cognition and health* (pp. 629–649). John Wiley & Sons.

Vane-Wright, R. I. (2022). Turning biology to life: Some reflections. *Biological Journal of the Linnean Society*, 1–18. https://doi.org/10.1093/BIOLINNEAN/BLAC141.

Voigt, C. A., Kauffman, S., Wang, Z. G. (2000). Rational evolutionary design: The theory of in vitro protein evolution. *Advances in Protein Chemistry, 55*, 79–160.
Wiener, N. (1948). Cybernetics. *Cybernetics or Control and Communication in the animal and the machine*. The MIT. https://doi.org/10.7551/MITPRESS/11810.001.0001.
Williams, D. M., & Ebach, M. C. (2020). Introduction: Carving Nature at its joints, or why birds are not dinosaurs and men are not apes. *Cladistics* (pp. 1–10). Cambridge University Press. https://doi.org/10.1017/9781139047678.002.
Woods, R. J., Barrick, J. E., Cooper, T. F., Shrestha, U., Kauth, M. R., Lenski, R. E. (2011). Second-order selection for evolvability in a large Escherichia coli population. *Science, 331*, 1433–1436.

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