Biology & Philosophy (2021) 36:8
https://doi.org/10.1007/s10539-021-09780-8

A universal ethology challenge to the free energy principle: species of inference and good regulators

Michael D. Kirchhoff[1] **· Thomas van Es**[2]

Received: 30 July 2020 / Accepted: 22 January 2021 / Published online: 18 February 2021

Abstract

The free energy principle (FEP) portends to provide a unifying principle for the biological and cognitive sciences. It states that for a system to maintain non-equilibrium steady-state with its environment it must minimise its (information-theoretic) free energy. Under the FEP, to minimise free energy is equivalent to engaging in approximate Bayesian inference. According to the FEP, therefore, inference is at the explanatory base of biology and cognition. In this paper, we discuss a specific challenge to this inferential formulation of adaptive self-organisation. We call it the *universal ethology challenge*: it states that the FEP cannot unify biology and cognition, for life itself (or adaptive self-organisation) does not require inferential routines to select adaptive solutions to environmental pressures (as mandated by the FEP). We show that it is possible to overcome the universal ethology challenge by providing a cautious and exploratory treatment of inference under the FEP. We conclude that there are good reasons for thinking that the FEP can unify biology and cognition under the notion of approximate Bayesian inference, even if further challenges must be addressed to properly draw such a conclusion.

Keywords Free energy principle · Biology · Cognition · The universal ethology challenge · Unification · Inference · Models

✉ Michael D. Kirchhoff
kirchhof@uow.edu.au

[1] School of Liberal Arts, Faculty of Arts, Social Sciences and Humanities, University of Wollongong, Wollongong, Australia

[2] Centre for Philosophical Psychology, Department of Philosophy, Universiteit Antwerpen, Antwerp, Belgium

Introduction

Background

The concept of inference is taking on a strikingly prominent role in contemporary cognitive neuroscience, theoretical biology and philosophy of cognitive science. This is especially clear within Bayesian models of biology and cognition such as predictive coding (Rao and Ballard 1999), predictive processing (Clark 2013; Hohwy 2016), the Bayesian brain (Knill and Pouget 2004), and the Cybernetic brain (Seth 2015). These Bayesian models are effectively theories of the structure, function and dynamics of the brain (Kirchhoff and Froese 2017; Ramstead et al. 2019). Although formulated within a Bayesian framework, the *free energy principle* (FEP) is a much broader reaching framework, seeking to provide a general theory unifying biology and cognition—formulated almost entirely from mathematical principles in physics and machine learning (see e.g., Friston 2010, 2013; Hohwy 2013; Kirchhoff et al. 2018; Linson et al. 2018; see also the introduction to this special issue). Its ambition is to secure a formulation of systems that are in non-equilibrium steady-state with their environment by appealing to formalisms in machine learning, information and probability theory, and then employ those formalisms to derive an explanation of biological self-organisation and cognition within the same framework (Friston 2019; Hesp et al. 2019). In this paper, our focus will be exclusively on the FEP.

What does a free energy treatment of biology and cognition look like? It would be an account of biology and cognition based on approximate Bayesian inference and the notion that an organism or phenotype can be understood as a statistical model of its environment. The FEP suggests that all organisms are driven to minimise an information-theoretic quantity known as 'free energy'. Mathematically, free energy is a bound on surprise. Surprise is an improbability measure of some outcome (Corcoran et al. 2020; this issue), where the outcome in question (typically) refers to some sensory state dependent on the relation between action and external causes of sensory states. Under the FEP, negative surprise is equivalent to Bayesian model evidence. This means that a system (e.g., a cell) that minimises free energy is a system that accumulates evidence for its own model of how its sensory input was generated. In machine learning, this process is the same as approximate Bayesian inference. Organisms that succeed in minimising free energy, the FEP says, do so by approximating Bayesian inference. In short, the FEP provides an inferential and statistical view of biology and cognition (cf. Schrödinger 1944).

Aim and argument

The FEP is both exciting and controversial. A central worry is that the FEP places *inference* at the explanatory base of adaptive self-organisation (or life) and cognition, including action and perception. Some suggest that the application of inference to biology is a serious error, illustrating an anthropogenic bias in explanations

of biology and cognition (Lyon and Keijzer 2007). Others think that such a liberal application of the concept of inference renders it insubstantial as an explanatory concept (Orlandi 2017). Inference may be involved in more sophisticated forms of cognitive processing such as counterfactual reasoning, or thinking, but not for more basic forms of early visual processing (in perception) or chemotaxis (in metabolism). Others still have argued that the suggestion that life or adaptive self-organisation is inferential conflates evolutionary facts about lineages with proximate facts about individuals, given that it is far from obvious that inference (or prediction) is required for life (Sterelny 2005).[1]

In this paper, we focus on the last articulation of this worry about inference, which we label the *universal ethology challenge*: it states that the FEP cannot unify biology and cognition, for life (or adaptive self-organisation) itself does not require that organisms minimise free energy via approximate Bayesian inference.[2] Our aim will be to show that it is possible to overcome the universal ethology challenge, which will, by extension, establish that inference under the FEP is not explanatorily weak. We will do so by arguing that approximate Bayesian inference can be shown to be involved over a continuum of biological processes, from early visual processing in perception to chemotaxis in bacteria.[3] In all of these cases, approximate Bayesian inference can be understood without organisms having explicit knowledge of the prior probability distributions over environmental causes of their sensations. We shall conclude that there is no substantial difference between perceptual processing in animals and chemotaxis in bacteria—these different adaptive dynamics all conform with approximate Bayesian inference.

One immediate question that our argument faces is an issue about the ontological status of inference under the FEP. On the one hand, it is possible to argue that the FEP yields a picture of biology and cognition as inherently inferential and statistical. This would be a *metaphysical* (realist) interpretation of inference under the FEP. More common is the *methodological* approach, which models biological and cognitive characteristics by using concepts from information theory and Bayesian statistics.[4] All that will be important for our argument is that biology and cognition can be modelled as if their characteristics can be captured in inferential and statistical terms. There are, we think, two general points that lend support to this

[1] Sterelny (2005) does not target the FEP. His target is niche construction theory. Yet, we suspect his worries carry across to the FEP. We leverage his objection to inference or prediction as an inherent feature of biological self-organisation in Sect. 3 of this paper, under the notion of the universal ethology challenge.

[2] We add 'adaptive' to self-organisation here because there are plenty of examples of self-organising systems that we would not want to say are alive. For example, hurricanes are self-organising phenomena. Yet they are not (or so we shall assume) alive. Rocks are self-organising systems too. But rocks are not living systems. This suggests that the ability to actively modulate one's relation to environmental perturbations is a core feature of life. We employ the concept of adaptive self-organisation to capture this *active* aspect of living systems (see also Godfrey-Smith 2018).

[3] Although the inclusion of additional phenomena would be nice, it is not plausible given restrictions on the length of papers for this special topical issue.

[4] There is now a growing discussion of these issues within the philosophy of the FEP, and Bayesian approaches to cognitive science more generally (Colombo et al. 2018).

methodological option. One is a *meta-theoretical* point in the sense that adopting the FEP creates a common formal language within which different research communities can approach biological and cognitive phenomena.[5] There is also a *theoretical* point supporting this: it is possible to capture biological and cognitive characteristics in the formalisms underwriting the FEP. Indeed, the FEP is intended to apply to any system able to maintain its organisation despite tendencies towards disorder: from chemotaxis in cells (Friston 2013; Auletta 2013), neuronal signalling in brains (Friston et al. 2017; Parr and Friston 2019), tropism in plants (Calvo and Friston 2017), synchronised singing in birds (Frith and Friston 2015) to decision-making and planning in mammals (Friston 2013; Williams 2018). It has also been applied to model adaptive fitness over evolutionary timescales by casting evolution in terms of Bayesian model optimisation (Hesp et al. 2019).[6]

Another suspicion some readers might harbour is that Bayesian inference in the FEP is not actually Bayesian inference. There is some truth to this suspicion. Even if it is possible to cast biological and cognitive phenomena through the theoretical lens of Bayesian inference, it does not follow that organisms should be modeled as performing optimal or exact Bayesian inference. Under the FEP, variational inference is an *approximation* to exact Bayesian inference (Wiese 2017). In exact Bayesian inference the posterior probability of the causes of data conditioned on the data is treated as tractable in light of Bayes' theorem (see Sect. 2.3). In the FEP, to minimise free energy is to assume a posterior distribution that functions as a proxy for the actual posterior, given that organisms cannot know the causes of their sensory input (see Sect. 2.2). Broadly speaking, the FEP states that adaptive self-organisation consists

[5] We acknowledge that this is not an uncontroversial point. For example, those influenced by the Levins framework (see e.g., Weisberg 2006 in this journal) might object that seeking to model diverse phenomena from physics to biology and cognitive science within a single common language necessarily implies a problematic trade-off between generatility and precision, on the one hand, and generality and accuracy (i.e., realism), on the other. That is, the FEP has the look and feel of a framework that illustrates Levins' (1966) second strategy to idealised model building, which sacrifices realism for generality (i.e., seeking to unify explanations of as many phenomena as possible) and precision (i.e., the fineness of specification of parameters, variables, and other parts of model descriptions allowing for specific mathematical modelling). As Weisberg puts it: "Generality and realism however are important for a much more fundamental goal of scientific inquiry, giving scientific explanations." (2006, p. 640) We note this point of controversy here, only to set it aside. It is a serious issue, but here we will leave it an open question whether the FEP necessarily trades off realism (in the explanatory sense of the concept) for generality and precision. Settling this issue is a task for a different paper. We would like to thank Ross Pain for bringing this point to our attention.

[6] Following on from the previous footnote. It could be argued that this extreme scale of generality is a vice, not a virtue of the FEP. We address this issue elsewhere: Van Es & Kirchhoff (under review), 'between pebbles and organisms: weaving autonomy into the Markov blanket' *Synthese*. This issue has also received attention in Kirchhoff (2018). A further point brings us back to whether it is problematic to model all dynamical systems with the same and single formal framework. This tension here is between lumping explanations of diverse phenomena into only one explanatory framework or whether several different yet complementary frameworks are required to gain this degree of explanatory generality. Our concern here is not to settle this specific issue. We focus only on the question of whether inference is at the explanatory base of adaptive self-organisation and cognition. For initial discussion, see Ramstead et al. (2017).

in organisms seeking to recover a distribution, q, such that this distribution will be a good enough approximation of the actual posterior, p.

A bit more clarification and precision is needed. The FEP conceives of all biological organisms as engaged in one and the same activity; namely, free energy minimisation qua approximate Bayesian inference. Here 'inference' is often understood in terms of *counterfactual* inference realised in hierarchical models. Counterfactual inference captures the important ability of organisms to anticipate the kind of sensations they would expect to encounter were they to undertake certain actions (e.g., one might expect increases in heart rate given running on the beach). This orientation towards minimising surprise over time (where the expected surprise over time is known as entropy), implies the presence of models with temporal depth, i.e., models able to generate inferences or predictions about temporally nested causes of sensations (Parr et al. 2018). Here we adopt the proposition that model optimisation (or, conversely, surprise minimisation) is diachronic (Kirchhoff 2015; Kirchhoff and Kiverstein 2019a) and that this implies temporally deep models (Corcoran et al. 2020, this issue). The important thing for us is that the notion of temporal depth is one of degree. Some organisms (humans, for example) will exhibit thicker and deeper processing architectures. Others (e.g., plants or bacteria) have thinner and shallower ones. This difference in temporal depth does not mark a difference in the way organisms can be cast as inferring the causes of their sensations. Rather, it suggests that different organisms can be modelled as having different *degrees of freedom*—i.e., different ways in which they can maintain their internal stability in light of pressures from the environment.[7]

Our plan is as follows. We start by rehearsing the basic tenets of the FEP; how it derives a Markov blanket conception of organisms; and provides a formal provides a formal treatment of how to minimise surprise; and factorises into two notations for free energy minimisation—perceptual and active inference (Sect. 2). We will need to spend some time introducing these aspects, as they will be important for how we will address the universal ethology challenge. We proceed to develop the universal ethology challenge against the FEP (Sect. 3). We then consider how inference under the FEP can be shown to be involved across a continuum of biological processes, from early visual processing to chemotaxis in bacteria. We conclude that this provides good reasons for thinking that the FEP can unify biology and cognition under the notion of inference, even if further challenges must be addressed to properly draw such a conclusion (Sect. 5).[8]

[7] The term 'degrees of freedom' is a statistical notion. It speaks to the idea that different systems may have a number of different ways in which they can move without violating essential constraints imposed on them. It is these different ways of moving that are referred to as 'degrees of freedom'.

[8] We cannot hope to canvas enough examples to conclude, once and for all, that explaining any kind of adaptive behavior in terms of approximate Bayesian inference is required, even under the FEP. This naturally implies that the burden of proof for our main conclusion cannot be wholly and exhaustively provided. We acknowledge this, and hope that by focusing on very basic or simple kinds of activity, we provide enough evidence to motivate our main argument as a legitimate possibility. However, there is a reason to believe that we are on the right track. This reason is that inference under the FEP implies that organisms work to remain within nonequilibrium steady state, i.e., to maintain their phenotypic form (Kirchhoff & Kiverstein 2019b; Ramstead et al. 2020). We thank an anonymous reviewer for this observation.

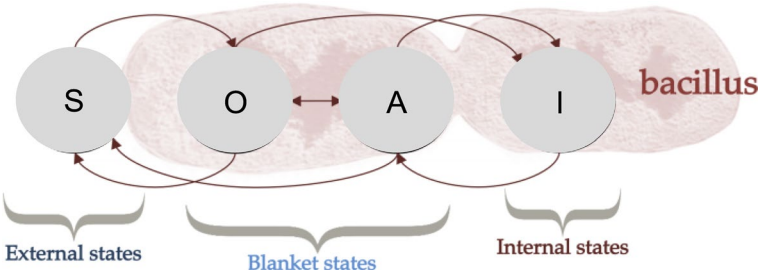

Fig. 1 The Markov Blanket and its partitioning into internal and external states separated by sensory, *o*, and active states, *a* (adapted from Friston 2019, p. 8)

The free energy principle

The Markov blanket formulation: nonequilibrium, ergodicity and shannon entropy

The FEP specifies what characteristics a system must exhibit for it to exist (Friston 2013; Hohwy 2020). It starts from the observation that "any […] random dynamical system that possesses a Markov blanket will appear to actively maintain its structural and dynamical integrity." (Friston 2013, p. 2) The notion of a Markov blanket defines the boundaries of a system (e.g. a cell or a multicellular organism) in a statistical sense. It is a statistical partitioning of a system into internal states and external states, where the blanket itself consists of the states that separate the two. The states that constitute the Markov blanket can be further partitioned into active and sensory states (Fig. 1).

The cell is an intuitive example of a living system with a Markov blanket. Without possessing a Markov blanket a cell would no longer be, as there would be no way by which to distinguish it from everything else (Palacios et al. 2020). This means that if the Markov blanket of a cell deteriorates there will be no evidence for its existence (statistically, speaking), and it will cease to exist (physically, speaking). Evidence for the existence of any system at non-equilibrium steady-state with its environment is thus premised on it having a Markov blanket (Friston 2013; Kirchhoff et al. 2018; Ramstead et al. 2017).

The Markov blanket formulation of biological organisms suggests that for organisms to maintain their integrity (of their internal and blanketed states) they must on average minimise the dispersion of their sensory states, *o*, which they can do via changes in their active states, *a*, impacting on the external states, *s*, causing sensations. This conditional dependency of sensory input upon action is represented by Friston and Stephan (2007, p. 424) as follows: $p(\tilde{o}) \rightarrow p(\tilde{o}|a)$.[9] Put differently, the

[9] Note that 'states' denote any variable that locates the system in question at some particular time in its state space. In the case of am embodied central nervous system, say, active states refer to the set of all actuators or effectors, and sensory states to the states of all sensory organs. Furthermore, as one of the reviewers point out, the notation here expresses the entailment relation between two probability densities; hence, not the relation between states as such. However, like Friston & Stephan (2007), all we wish to say here is that sensory data or outcomes are dependent on action.

Markov blanket for any system can be defined by the absence of connectivity; that is, internal states do not cause sensory states, and external states do not cause active states. From this it is possible to unpack the notion that living systems self-organise to a set of attracting states far from thermodynamic equilibrium—this is the idea (under the FEP) that any living system is endowed with a particular density, a non-equilibrium steady state (NESS) density. It is the formulation of the FEP as based on a NESS density that allows for the related formulation that organisms are locally *ergodic*. The better organisms are at reducing the dispersion of their sensory states the more likely it is that they will occupy a limited set of sensory states. In other words, one can expect biological systems to frequent or be in a relatively small set of attracting states—the states that comprise their phenotype (Corcoran et al. 2020, this issue; Friston 2013). In probability theory, this is the same as saying that the average amount of time a state is occupied (e.g., a grizzly bear hunting fish) is equal to the probability of the system being in that state when observed at random. Hypothetically, we could imagine that a grizzly bear spends 0.5 of its time awake hunting for fish. This implies that the conditional probability of observing the bear hunting for fish when observed at random during the time it is awake would be 0.5.[10]

The existence of an attracting set of states (defining an organism's phenotype) implies that the probability distribution over an organism's (interoceptive and exteroceptive) sensory states must have low entropy (Friston 2011)—there is a set of states it is required to frequent to remain in its same phenotypic form. Assuming that the grizzly bear is either in one state (catching fish) or in another state (sleeping), the probability distribution when observed catching fish or sleeping would have low surprise. However, a fish out of its aquatic milieu would be in a state with high surprise (conditioned on its phenotype).

Entropy is understood information-theoretically in the FEP. In information theory, Shannon entropy is the weighted average of surprise (or self-information). Self-information is the negative log-probability (surprise) of something happening (e.g., it would be surprising to be falling off the face of a cliff-edge when riding a mountain bike). Note that entropy is the average surprise associated with some event or outcome. Parr and Friston (2017) formalise the average Shannon entropy of the dispersion of outcomes in the following way: $H[P(\tilde{o})] = -E_{p(\tilde{o})}\ [\ln P(\tilde{o})]$. $H[P(\tilde{o})]$ denotes the Shannon entropy of a distribution over outcomes; $\tilde{o}$ means a sequence of sensory outcomes ($o_1, o_2, o_3, \ldots, o_n$); and $-E_{p(\tilde{o})}$ refers to the surprise or negative expectation with respect to $\tilde{o}$. To exist therefore entails minimising surprise over a distribution of sensory states (under the FEP).

Bounding surprise and tractability issues

In the previous section we portrayed the FEP's starting point; namely, for an organism to remain in non-equilibrium steady-state with its environment it must maintain

[10] A different example would be tossing a fair coin, where the average amount of time a fair coin lands on 'heads' is equal to the probability of the system (coin) being in that state (heads) when observed at random. Since the coin toss is fair, the average time is 0.5 (if the timescale is 0–1) and the probability is 0.5 (as we measure probabilistic outcomes on a scale from 0–1).

its sensations within a specific range (e.g., maintaining homeostasis via allostasis (Seth 2015; cf. Corcoran et al. 2020, this issue)). Organisms must minimise the dispersal (surprise) associated with their sensory states. The key question is: How do they do it?

The problem with surprise minimisation is: a system "cannot know whether its sensations are in fact surprising" (Friston 2010, p. 128). To see this, consider that surprise minimisation is the same as maximising or optimising the following quantity: $Q(\tilde{o}|a) = \ln p(\tilde{o}|a,m)$. Under the FEP, for an organism to exist it must on average take measures to optimise $Q(\tilde{o}|a)$. This is simply because the negative log probability of $p(\tilde{o}|a,m)$ is a measure of the surprise associated with the dispersion of sensory states conditioned on action and a model of how the sensory states were generated (Friston and Stephan 2007). Crucially, in statistics $Q(\tilde{o}|a)$ is known as the marginal likelihood or *model evidence*. This is important, for it implies that acting to minimise surprise is equivalent to maximising evidence for one's model (Hohwy 2020). This process is known as *self-evidencing* (Hohwy 2016). Direct evaluation of the surprise of sensory states is however computationally intractable. The reason is that for any given organism, its actions must optimise the following notation: $Q(\tilde{o}|a) = \ln \int p(\tilde{o},s|a)ds$ (Friston and Stephan 2007, p. 425). Here $p(\tilde{o},s|a,)$ is the joint probability of sensory outcomes and their (hidden) external causes conditioned on action. Yet it is not possible to compute the integration over hidden (external) states to define the joint distribution, $p(\tilde{o},s)$. This follows from the Markov partitioning rule, which renders external states hidden from internal states given sensory and active states. So an organism does not have direct access to the causes of its sensations.

The FEP is a formal demonstration of how to overcome this intractability problem (i.e., the problem of not knowing one's model). It works as follows: (1) Free energy, F, is a function of the states organisms can access; namely, their sensory and internal states separated from external states by the presence of a Markov blanket; (2) By definition, F is $\geq$—$\ln p(\tilde{o}|a,m)$, which implies that F is always equal to or greater than surprise; and (3) given that the time average of surprise is the same as entropy, the FEP solves the problem of how an organism could in principle slow down the inevitable effects of the second law of thermodynamics.

Species of free energy and approximate inference

The FEP proposes that any system with a Markov blanket can be "interpreted as embodying a process of … inference which minimizes a single information-theoretic objection—the variational free-energy." (Millidge et al. 2020, p. 1) This is colloquially understood as *perception* (Parr et al. 2018). *Action* becomes part of this formulation under the notion of *active inference* (Friston 2010; Millidge et al. 2020). Since active inference usually targets temporally extended action sequences, sequences of actions that minimise the sum of free energy over time are captured by a quantity known as *expected free energy*. This means that the free energy quantity, F, can be defined both in terms of a *variational free energy*, *VFE*, and an *expected* free energy, *EFE*. In the remainder of this section, we introduce variational free energy for action

and perception without unpacking its future-oriented, expected form. We return to EFE in Sect. 4.2.[11]

Friston and Stephan (2007) define variational free energy as comprised of two densities: the *recognition* (or ensemble) *density*, $q(s;i)$, and the *generative density*, $p(\tilde{o},s|a)$, respectively. A generative density is comprised of a *prior*, $p(s)$, and a likelihood function, $p(\tilde{o}|s,a)$, making up a generative model for any system, $p(\tilde{o}|s,a)p(s)$.[12] A generative model is a statistical mapping from sensory outcomes to external (hidden) causes given prior beliefs about the causes and a likelihood function determining the generation of sensory input given external causes (Parr et al. 2018).[13] Generative models thus describe the process by which sensory outcomes are (expected) to be generated. The ensemble density is an approximate posterior density over external (hidden) causes of sensations (Hohwy 2020)—i.e., it is a density over the causes of the generative model. This means that organisms can be interpreted as instantiating a generative model from which to infer the causes of sensations; a process which can be used to train (update) the ensemble density—the organism's posterior expectations about the causes of outcomes (cf. Dayan et al. 1994).[14] This has a clear adaptive function: it is important to know whether the cause of one's sensory input is a fast approaching grizzly bear or something quite different, and harmless. Formally, the variational free energy *for perception* is expressed in the following equation (Friston & Stephan 2007, p. 427):

$$F = -\ln p(\tilde{o}|a) + D(q(s;i) \parallel p(s|\tilde{o},a)).$$[15]

Though the mathematical reasoning behind this is complicated, the basic idea is that by minimising free energy an organism can be viewed as tightening the Kullback–Leibler (KL) divergence between the *recognition* (ensemble) *density*, $q(s;i)$, and the conditional density of the causes of sensory outcomes, $p(s|\tilde{o},a)$. This has the additional (interesting) implication that any system that minimises its free energy via perceptual inference can be cast as engaging in approximate Bayesian inference—as *inverting* the generative model to give the recognition density (Ramstead et al. 2019). To see why this is the case consider that when the VFE for perception is minimised, the ensemble density becomes approximately equal to the true posterior. This is the same quantity that Bayesian inference seeks to optimise, via Bayes' theorem:

[11] We wish to thank one of the reviewers for enabling us to make this point more precise, ensuring that the discussion (in more informal terms) of expected free energy is removed from this section of the paper.

[12] A 'prior' is the probability of a belief or an expectation about hidden causes of sensory outcomes independent of sensory outcomes. A 'likelihood' is a distribution that determines how hidden states cause sensations.

[13] A different way of putting this is: a generative model is a joint probability density over all the states that comprises the system; namely, internal states, blanket states, and external states. It is these joint probability densities which can be described as a set of priors and likelihoods.

[14] See Ramstead et al. (2019) for a more detailed account of the relationship and characteristics of these two densities.

[15] Here we again slightly change the terms used in this equality for consistency purposes only.

$$(A) : p(s|\tilde{o}) = \frac{p(\tilde{o}|s)p(s)}{p(\tilde{o})} \, (B) : p(s|\tilde{o}, a) \cong \frac{p(\tilde{o}|s,a)p(s)}{p(\tilde{o}|a)}$$

In (A) we provide the standard notation for Bayes' theorem. Bayesian inference under Bayes' theorem is said to be *optimal* if and only if: (a) the prior is chosen appropriately; and (b) the posterior is tractable (i.e., the causes of outcomes can be directly inferred by using the computations on the right side of this equality). Setting (a) aside, (b) is generally not true when causes and outcomes are non-linear. In this specific sense, then, the FEP is not the claim that organisms perform optimal or literal Bayesian inference to minimise surprise. In (B) we capture this by highlighting that free energy minimisation conforms to *approximate* Bayesian inference. In the equation above, this involves assuming a posterior that functions as a *proxy* for the actual (or true) posterior distributions. Approximations do not guarantee a tight fit between the distributions. This is why minimising VFE via perceptual inference results from work done by the organism to update its best 'guesses' about the causes of sensations in light of new incoming sensory evidence. On average, the FEP implies, this should minimise the KL- divergence between the approximate posterior and the actual posterior, ensuring a tight fit or coupling between what organisms expect to be the case and what is actually the case.

Bounding surprise in this statistical fashion also turns on *active states*. Given that organisms must minimise surprise to maintain their structural and functional integrity, the FEP states that organisms must select policies (i.e., action sequences) they expect will result in the lowest free energy, on average and over time. This is active inference, where free energy is reduced with respect to sensory outcomes dependent on the conditional dependency between action and causes of sensory outcomes. Friston & Stephan (2007, p. 427) define this in the following notation:

$$F = -\ \langle \ln p(\tilde{o}|s,a) \rangle_{q} + D(q(s) \parallel p(s)).$$

The first term refers to the surprise associated with sensory outcomes.[16] The second term expresses the KL-divergence between the inferred (posterior) causes of sensory states and the actual causes. Whereas it is the KL-divergence that is minimised in perceptual inference, it is the first term that is minimised in active inference. This is because it is the only term that itself is a function of action. Crucially, minimising free energy via action corresponds to optimising the accuracy of sensory outcomes.

Two things are especially important to note. First, barring pathologies to act otherwise (i.e., self-harm or suicide), any system that seeks to reduce its free energy (for adaptive purposes) will select sequences of action that expose it to the kind of sensations it expects (e.g., if I expect to be hungry I infer the action policies that, if selected, would elicit my expected sensations—being full). Second, systems that occupy a space with a low free energy minima will be systems that (on average) visit and revisit their expected states, where the expectations specify their ensemble density (this follows from the ergodicity assumption). Finally, inferring policies

[16] Formally, it is known as the negative expectation under the log probability of sensory outcomes conditioned on the joint distribution between external (hidden) states and active states.

resulting in low-free energy outcomes ensures that the KL-divergence between the inferred causes approximate the actual causes of sensations remains tight. Within the FEP, perception and action therefore are two co-dependent aspects of precisely the same imperative to reduce free energy (Kirchhoff and Robertson 2018; Millidge et al. 2020; Ramstead et al. 2019).

The universal ethology challenge

One immediate and difficult question about non-equilibrium steady-state systems is their *inferential* formulation under the FEP. This formulation has been challenged. Orlandi (2017) suggests that such a formulation is not a substantive thesis, for it is not "substantive to say that a system is inferential simply because it can be described as performing inferences." (2017, p. 18) Many things can be interpreted as if they perform inferences—e.g., a pair of coupled pendulums (Friston 2013). Clark (2017) draws a similar conclusion as he considers single-celled organisms capable of survival-enhancing chemotaxis. He says that such "a life-form may respond to environmental perturbations using a variety of tricks and ploys, none of which require it to engage in a process in which incoming sensory stimulations are met with attempts to generate the incoming signal 'from the top down' using stored knowledge about the world." (2017, p. 4) He concludes that "talk of such a being 'predicting' such-and-such … is either simply false or merely short-hand for what is really a rather different claim … To describe this whole simple (reactive, feed-forward) creature as a 'model' of its world … can also seem somewhat strained." (2017, p. 4).

In a different context, Sterelny (2005) objects to Odling-Smee et al's (2003) thermodynamic treatment of niche construction theory on similar grounds.[17] In this section, our agenda will be to leverage Sterelny's critique of niche construction as a *universal ethology*, because this critique can be shown to map directly onto the FEP—which itself can be understood as advancing a universal ethology; a set of features or characteristics all living organisms must have if they are to exist. According to Sterelny, the specific rendition of niche construction theory by Odling-Smee et al. (2003) states that all living systems must be niche constructors in virtue of being far-from-equilibrium systems. That is, organisms are energy- and-entropy pumps: "they pump energy from the environment, and pump entropy into it. These thermodynamic preconditions of life define a universal ethology: a set of characteristics all living agents must have." (Sterelny 2005, p. 24) The specific characteristics are the following:

1. Organisms must be *active*: organisms need to undertake certain activities to secure the energy resources required for existence;

[17] According to Laland et al. (2016), niche construction is (broadly defined) the process by which organisms modify their own evolutionary niches, including the niches of others (Odling-Smee et al. 2003). When these modifications change natural selection pressures, evolution by niche construction is a possible outcome (Laland et al. 2016).

2. Niche construction requires energy, so it must on average be *profitable*;
3. Niche construction involves *discrimination*. Since environments and organisms vary, the kind of niche constructing that is profitable will vary too. This implies that an organism's niche-constructing behavior must be controlled by systems well-designed for their local environment; and finally:
4. Niche construction is *predictive*: given that niche construction is active, and actions unfold prior to immediate feedback from the environment, it follows that organisms must act on the basis of search plans. Hence, niche construction must be predictive or inferential in some specific sense.

Sterelny's assessment is that these preconditions for life—this universal ethology—do not hold. He says: "I am unconvinced that these thermodynamic conditions define a universal ethology. Life itself does not require active, future-oriented search by individual agents." (2005, p. 25) Consequently, Sterelny thinks the fourth condition—that niche constructing behavior must be predictive or inferential—does not hold for all living things. This is exactly the kind of conclusion that Clark arrives at when considering chemotaxis in *Escherichia coli* (*E. coli*)—there is no need to think that this adaptive process involves inference; and there is no need to think that single-celled organisms are (statistical) models of their world. The FEP however implies that any system that exists in a non-equilibrium steady-state with its environment must minimise its free energy and will therefore be a model of its (internal and external) environment. This follows from the observation that adaptive fitness and negative free energy is the same thing (Friston et al. 2012).

Transposed to the FEP, it is possible to derive Sterelny's conclusion for niche construction theory as holding with respect to the FEP. There are several reasons for why drawing this analogy is important and can be motivated. First, both frameworks provide a thermodynamic treatment of adaptive self-organisation; hence, they both begin from broadly the same set of background assumptions about life. Second, there is now work to suggest that the niche construction theory can be incorporated within the FEP (Constant et al. 2020); hence, if there is a problem with the former, then incorporating it within the latter will likely carry the same problem across. Specifically, we draw the analogy precisely to highlight that *if* Sterelny is correct in his assessment of the niche construction theory—that not all forms of adaptive self- organisation requires predictive activity on the basis of search plans—*then* the exact same problem arises for the FEP. The reason is that the FEP requires that all forms of adaptive self- organisation are explained by appeal to predictive or inferential activity on the basis of search plans (or, generative models). So, we draw this analogy not to question the first three pre- conditions comprising a universal ethology for life—i.e., activity, discrimination and profitability—but to focus attention on the inferential condition for life (according to the FEP).

In what follows, we derive Sterelny's conclusion for niche construction theory as holding with respect to the FEP by replacing 'niche construction' with 'free energy minimisation', accordingly:

1. Free energy minimisation is *active*: organism must undertake certain activities to remain alive, given that sensory outcomes are conditioned on action, $p(\tilde{o}) \rightarrow p(\tilde{o}|a)$.
2. Free energy minimisation requires energy [Gibbs free energy], so it must on average be *profitable*.
3. Free energy minimisation involves *discrimination*. Since environments and organisms vary, the kind of activities that are profitable will vary too. This implies that an organism's behavior must be controlled by systems well-designed for their local environment; and finally:
4. Free energy minimisation is *predictive* (inferential): free energy minimisation unfolds on the basis of probabilistic search plans (using the ensemble density to approximate the true posterior distribution). In active inference, organisms infer the action policies they predict will minimise expected surprise.

The problem is therefore clear. If inference is not required for life, it is not required for life. Or, differently put, if inference is not necessary for adaptive self-organisation (pace Clark 2017), it is not necessary for adaptive self-organisation. Should these observations hold, the FEP would not define a universal ethology. Consequently, the FEP cannot unify biology and cognition. It would not be sufficient, as it would be a mistake to say that all living things perform inferences to maintain their homeostatic variables within viable bounds (for example). According to Sterelny, to think life itself requires that organisms predict or infer causes of their sensations, is to "conflate evolutionary facts about lineages with proximate facts about individual organisms." (2005, p. 25) In defense of this claim, Sterelny considers a few different yet overlapping issues. Some plants can crank up production of defensive chemicals, others cannot (Sterelny 2005). The former are predictive (inferential) in an important sense (cf. Godfrey-Smith 1996). According to Sterelny: "Those with a fixed investment predict only at the level of the lineage, for individual response is determined by the level of threat registered by selection on the lineage as a whole. But individual agents in such lineages are not predicting the level of threat." (2005, p. 25) He makes a similar claim; this time with respect to filter-feeders. The ultimate causes of filter-feeding mechanisms are evolutionary, in the sense that there has been evolutionary sorting to select such mechanisms—but "the agent itself is not actively sorting." (Sterelny 2005, p. 25) He continues: "A filter-feeder has not even the most rudimentary search plan." (2005, p. 25) The problem that Sterelny articulates is a familiar one: if one extends a substantive notion sufficiently, it becomes too weak to capture anything substantive (in the explanatory sense of 'substantive').

Species of inference and good regulators

One size seldom fits us all. It is true of clothing. We shall now suggest that it is also true of inference. Before starting this part of our argument, we note that to the best of our knowledge there is no disagreement in the literature about the following: that adaptive self-organising activity is active, discriminatory and profitable (minimally

on average). The tension has rather to do with the issue of whether this kind of activity must be explained by appeal to inference or prediction on the basis of models or search plans. This is the reason for why we focus exclusively on the last condition of the universal ethology challenge.

Now, the fourth condition of the universal ethology challenge states, on the one hand, that organisms must act on the basis of search plans, and denies that this is true of all organisms, on the other. When articulated through the lens of the FEP, search plans take the form of generative models with various degrees of temporal depth. Given that adaptive self- organisation (under the FEP) implies that organisms minimise the free energy expected to follow from inferred actions, it follows that organisms must act on the basis of search plans (i.e., on the basis of generative models realised in action). The universal ethology challenge presents a problem for the FEP given that the FEP implies that all forms of generative models are search plans.

Our agenda in the next Sect. (4.1) will not be to prove the presence of search plans in adaptive self-organisation; it is rather to unpack how a minimal notion of inference as covariation is involved in early visual processing. In Sect. 4.2 we turn to finalise our assessment of the universal ethology challenge leveraged against the unification ambitions of the FEP. Here we focus on chemotaxis in bacteria and argue that even in this form of adaptive self-organisating behavior, it is possible to explain chemotaxis in terms of inference and search-plans. We end up arguing that it is possible to cast different cases from perception to chemotaxis in terms of inference, while making reference to certain positions skeptical of using terms such as inference and models to explain adaptive activity.

Inference in perception

A key aspect of generative models is hierarchy. In a hierarchical model the assumption is that hidden states generating sensory outcomes are themselves caused by hidden states at a higher scale of biological activity (this is usually illustrated by appeal to the laminar structure of cortical organisation). What is important for our purposes is the following: higher (cortical) areas respond to stimuli that change over longer timescales than lower areas. Or, differently put, dynamics at lower scales unfold faster than dynamics at higher scales. This feature is not specific to cortical organisation. It is a ubiquitous feature in biology, and in domains ranging from statistical thermodynamics, ecology to sociology (Kirchhoff and Kiverstein 2020). Swarm behavior in a flock of birds unfolds slower than the behavior of any individual bird. The rolling motion of fluid dynamics is slower than the molecular dynamics making up the ensemble behavior (Haken 1983). A sentence takes longer to read than a single word (Parr et al. 2018). Large-scale neuronal dynamics influence local neuronal behaviour by 'enslaving' local processing elements (Engel et al. 2001), and so on.

Under the FEP, these differences in timescale speak to the different degrees of temporality in generative models. Crucially, lower-level inference is therefore less temporally deep than inference at higher scales of the processing hierarchy, exhibiting highly specialised priors that guide inference about hidden states causing expected sensations. Girshick et al. (2011) report that human and nonhuman

animals *exploit* inhomogeneities in the orientation statistics of local environments and make use of these inhomogeneities in perception. There is evidence to suggest that neurons tuned to cardinal orientations of objects in natural scenes are over- represented in the primary visual cortex (Teufel and Fletcher 2020). These findings lend support to the idea that organisms harbour prior expectations about statistical regularities of environmental features reflected in early visual processing. The basic idea is that were one to draw a conclusion about population parameters (e.g., the orientation of trees) based on a sample taken from that population, the probability would be high that this sample would be over-represented with vertical orientations and only sparsely by horizontal orientations (fallen trees, say). Organisms living in such environments, especially upright and bipedal organisms, come to reflect such statistical relations in their physiology. The FEP describes this in terms of organisms instantiating prior beliefs about hidden causes of sensations; highly specialised priors over action policies guiding behavior.

As we know, the FEP states that organisms must infer the hidden states of the environment causing its sensory outcomes. The cardinal orientation of trees in a natural scene causes a pattern of photoreceptor activity (sensory outcomes) in the retina. Bayesian inference is used to approximate the probable orientation of objects (such as trees in a natural scene) from the available sensory data by utilising a highly specialised prior, $p(s)$, and a likelihood function, $p(o|s)$. Utilising prior expectations about cardinal orientations in guiding action will likely be less computationally and metabolically costly than engaging in more exploratory modes of inference.

We have chosen this case as our first case of inference because it seems so unlikely that there is anything like inference going on here. The finding that specific neurons in the primary visual cortex are tuned to cardinal orientations of objects signals the effects of evolutionary tracks laid down in organisms over time. It need not suggest that this has anything to do with inference or that it need be explained by appeal to inference.

We now illustrate that there is a straightforward way of understanding why the FEP implies an inferential interpretation of the relationship between neurons tuned to cardinal orientations of natural scenes and the actual scene statistics. We do this by unpacking this figure from Teufel and Fletcher (2020, p. 236):

This highlights the correlational relationship between prior distributions about orientations of objects in natural scenes and the actual environmental distribution. The correlation between the prior and the environmental regularity in this figure can be cast as approximate Bayesian inference. In the FEP formulation of perception, perceiving involves inferring the posterior probability, $p(s|o)$. In this case, this is the probability of the causes of outcomes (cardinal orientations of trees in a natural scene) conditioned on sensory outcomes (photoreceptor activity). This is achieved by inferring from observations to the causes of outcomes about which the organism has prior beliefs. In this case, priors correspond to distributions shaped by physical states—e.g., the specific configuration of orientation-tuned neurons in V1.

This maps onto the formalism for variational free energy, VFE, as follows: instead of trying to infer the causes of outcomes directly, the FEP says (as we know) that an organism assumes.

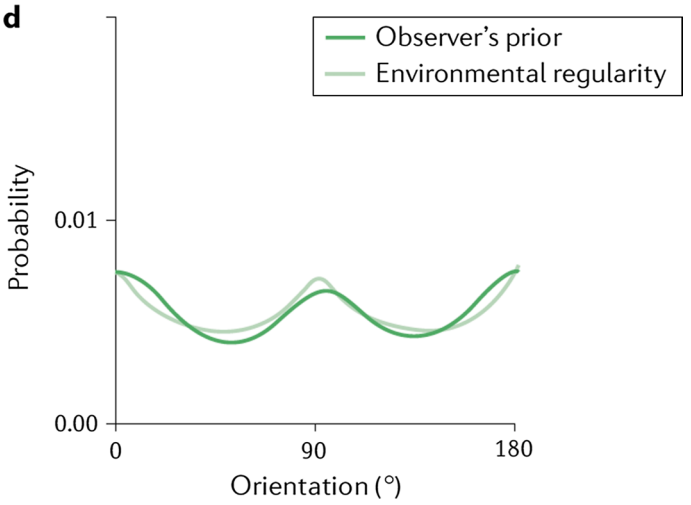

Fig. 2 The human visual system described as using a prior when judging local orientation has been derived from psychophysical data using a Bayesian framework. This prior (dark green line) shows a close correspondence to environmental regularities (light green line) (Teufel and Fletcher 2020, p. 236)

an ensemble density or recognition model, $q(s;i)$, which is used to approximate the actual or true posterior distribution, $p(s|o)$. Thus formulated, we can note that the distance between q and p is expressed in the KL-divergence, $D(q(s;i) \parallel p(s|,o)$, which measures the relative (Shannon) entropy between the two probability distributions. As Hohwy (2020) puts it: the KL-divergence "tells us how good q is as a stand-in for what the system should infer about external states beyond its boundary, given its sensations." (2020, p. X) Crucially, the KL- divergence provides the mutual information between q and p, which shows how information about q reduces that Shannon entropy of the other distribution, p, and vice-versa. Formally, mutual information is written $I(s;o) = I(s) + I(o)—I(s,o) \geq 0$. This is exactly what we see in Fig. 2 (above), where there is a high correlation between the prior and the environmental regularity. This means that the prior expresses a lot of information about the environmental regularity, and vice-versa. This implies that the divergence between q and p is small, or close to zero, i.e., there is little to no surprise expressed in the KL-divergence. Thus, by associating the dynamics of neurons in early visual processing with prior beliefs about natural scene statistics, it follows that these dynamics can be associated with approximate Bayesian inference in the service of perception. We take this to be a minimal but important notion of inference within the FEP.

One suspicion some readers might have is that this form of inference is best seen as a form of non-accidental correlation between variables and events. This is likely to be what drives Orlandi (2017) but also Sterelny (2005) to suggest that such a notion of inference is too weak as an explanatory notion of interest. It is however important to be precise about the notion of inference under the FEP. First, there is a notion of inference in the FEP that is more closely aligned with the concept of correlation, in the sense that the values of one variable infers or predicts the values

of another variable. Second, there is a notion of inference in the FEP where VFE is minimised, which also underwrites the mere correlational notion of inference just mentioned. Finally, a third sense of inference is the counterfactual notion, which takes the form of the *expected free energy*, EFE. This involves inferences of hidden states, priors and state transitions. It is important not to conflate these different notions of inference. However, they are species of inference nonetheless.

Inference in the sense of correlation should not be problematic. Indeed, it directly reflects the KL-divergence utilised to define the free energy quantity, given that the KL-divergence just is a measure of *relative entropy* (i.e., Shannon entropy), which can be reformulated in terms of non-accidental correlations between variables (Kirchhoff and Robertson 2018). Note that for Shannon, something is a source of information if it has a number of different states that might come about on a particular occasion. As Godfrey-Smith (2007) observes, "any other variable carries information about the source if its state is correlated with that of the source. This is a matter of degree; a signal carries more information about a source if its state is a better predictor of the source, less information if it is a worse predictor." (2007, p. 106) Crucially, it is thus this notion of information that underwrites the claim that certain neurons in V1 can be interpreted as inferring information about cardinal orientations of trees in a natural scene. We conclude that this notion of information and consequently inference does not add new and problematic notions to the domain of biology and cognitive science. This mathematical notion of improbability is a legitimate notion in these areas of research.

Chemotaxis in *E. coli*: inference in action and good enough models

Although this view of inference is attractive (in our view), it faces an immediate worry. Recall that Clark (2017) draws a similar conclusion to Sterelny (2005); although he does so in the context of single-celled organisms capable of chemotaxis. Here Clark concludes, to repeat, that "talk of such a being 'predicting' [i.e., inferring] such-and-such … is either simply false or merely short-hand for what is really a rather different claim … To describe this whole simple (reactive, feed-forward) creature as a 'model' of its world … can also seem somewhat strained." (2017, p. 4) A different way of putting this worry would be to raise a question that lingers unanswered; namely, why should any organism seek to reduce free energy (or 'seek' anything at all) rather than simply do enough to get along?

Chemotaxis is a control mechanism for bacterial swimming. It rests on ongoing sensorimotor dynamics, tuning the bacteria to its chemical niche. It is a process many take to involve memory, for 'where' to turn at time *t2* is conditioned on the distribution of 'good' and 'bad' chemicals affecting sensory receptors and what was the case earlier at *t1* (say). If the conditions are favorable, the *E.coli* swims straight. Conversely, if the conditions are less than favorable, it initiates a form of tumbling behavior. Why assume this involves minimisation of free energy, and therefore approximate Bayesian inference, in addition to being active, profitable and discriminable?

Before looking closer at the claim that chemotaxis can be explained by appeal to approximate Bayesian inference, we briefly attend to why chemotaxis must be active, profitable as well as discriminable under the FEP. Chemotaxis is *movement-based*, and therefore an active pursuit on behalf of the living system to increase exposure to attractive substances. This can be given a more foundational characterisation in terms of the FEP. As we mentioned in section two, the FEP assumes that living systems self-organise toward a set of attracting states removed from thermodynamic equilibrium. This set of attracting states is known as the nonequilibrium steady state (NESS) density (Friston et al. 2020; Ramstead et al. 2020). It is this density that defines an organism's phenotype. To maintain homeostasis, *E.coli* must act to ensure that they 'consume' on average substances that they are attracted to given their phenotype. Without action, the system in question would simply self-organise to thermodynamic equilibrium (think of a candle flame or a snowflake as an example of this). This naturally suggests that chemotaxis must in the long-run be *profitable*, for unless a living system is able to return to its set of attractive states (its NESS density), it will cease to exist. A related way of putting this is in terms of the negative free energy of an action policy, because this quality can be partitioned into an extrinsic and an intrinsic value (Friston et al. 2015). Here minimising expected free energy is the same as optimising the expected utility of prior preferences (e.g., swimming toward attracting states), while reducing uncertainty about the causes of profitable outcomes is to be the same as optimising evidence for a model (i.e., a phenotype). This is in itself an interesting outcome, for it suggests that chemotaxis can be further broken down into exploitation activity (i.e., swimming) and exploration activity (i.e., tumbling). Note that striking the right balance between exploration and exploitation turns on the ability to *discriminate* on behalf of the organism. We consider this in more detail below, when we introduce the notion of temporally deep generative models.

We now turn to consider why chemotaxis can be explained by appeal to inference.

Survival-enhancing chemotaxis results in bacteria frequenting (on average) the kind of states (or chemical gradients) they expect to be in conditioned on their phenotype. If this were not the case, such bacteria would not exist. Mathematically, any system that does this will be a system whose dynamics can be understood as minimising free energy. That is, if a system minimises free energy, it will be a system that occupies its expected states on average. This means that if an organism is able to act in an adaptive way to fluctuations in its sensory states, it will minimally look as if it is seeking to reduce the expected surprise following its actions. As we know, in information theory, the expression—ln $p(\tilde{o}|a)$ is the surprise and it refers to the degree of uncertainty or unexpectedness of some event (under some action). It therefore represents the mismatch between what an organism expects given its action and what actually happens. A system that can minimise this mismatch will be a system that can tighten the KL-divergence or relative entropy over two probability distributions: the ensemble density and the conditional density of the causes of outcomes. In this sense, the chemotactic behavior of *E.coli* conforms to the formalisms of the FEP. This suggests that chemotaxis can be understood as involving prior beliefs about the kind of actions an organism can engage in, which, in turn, speaks

to the notion that *E.coli* must infer what course of action will result in preferred (metabolic) sensations in the immediate future (cf. Auletta 2013).

Clark (2017) finds this line of reasoning problematic, for he thinks that such "a life-form may respond to environmental perturbations using a variety of tricks and ploys, none of which require it to engage in a process in which incoming sensory stimulations are met with attempts to generate the incoming signal 'from the top down' using stored knowledge about the world. Such a being, though living and perfectly able to resist the second law by exchanging entropy with its environment, could be operating in a purely 'feed-for-ward' manner, responding to detected chemical gradients in ways not nuanced by any form of top-down predictive flow." (2017, p. 4) Moreover, he claims that to "describe this whole simple (reactive, feed-forward) creature as a 'model' of its world, though common in this literature, can also seem somewhat strained." (2017, p. 4).

Clark (2017) wants to accept the model-based and predictive account of cognition, including action and perception. Yet, at the same time, he stops short of applying this account of self- organising and adaptive activity in cases such as chemotaxis. If Clark is right about this, then schemes such as the FEP cannot reach all the way down to single-celled organisms such as *E.coli* simply because inference is at the explanatory base of all forms of adaptive self- organisation. What we will argue below is that this kind of model-free vision will not work as an explanation of chemotaxis. In other words, defenders of the FEP cannot both accept model and inference based explanations for some phenomena, and deny such explanations of other phenomena—on pain of theoretical inconsistency.

It might seem difficult to understand, or even counter-intuitive, to say that the dynamics exhibited by *E.coli* should be understood as approximate Bayesian inference conditioned on the bacteria being a model of its environment. Yet this follows naturally from the Good Regulator theorem in the field of cybernetics, given that Good Regulator theorem implies Bayesian model optimisation (cf. Conant and Ashby 1970). It states that a system (organism) is only able to regulate its relation to a larger system (environment) if it is a good model of that environment (Linson et al. 2018). A good regulator is therefore a system that can maintain its internal stability despite increasing (entropic) impacts from its larger environment. This means that *E.coli* can be said to be or become close-to-optimal models of their environment because surprise is defined as the negative log probability of sensory outcomes conditioned on a model (Friston et al. 2012).

Interestingly, a good regulator under the FEP is a model with a search plan in virtue of a generative model for minimising free energy. To see this, consider the difference between following two generative models:

The FEP implies that chemotaxis in *E. coli* presupposes (minimally) a generative model that takes the form of the left-side generative model in this figure. The transitions from s_1 to o_1 maps how hidden states generate sensory outcomes: relative concentration of attractants or repellants generate sensory receptor activity in *E. coli*. The inverse transition from o_1 to s_1 illustrates that hidden states are inferred from sensory outcomes based on a prior (D) about hidden states and a likelihood function (A) determining how hidden states generate sensory outcomes. The notation $B\pi_1$ is the beliefs an organism has about dynamically changing hidden states based

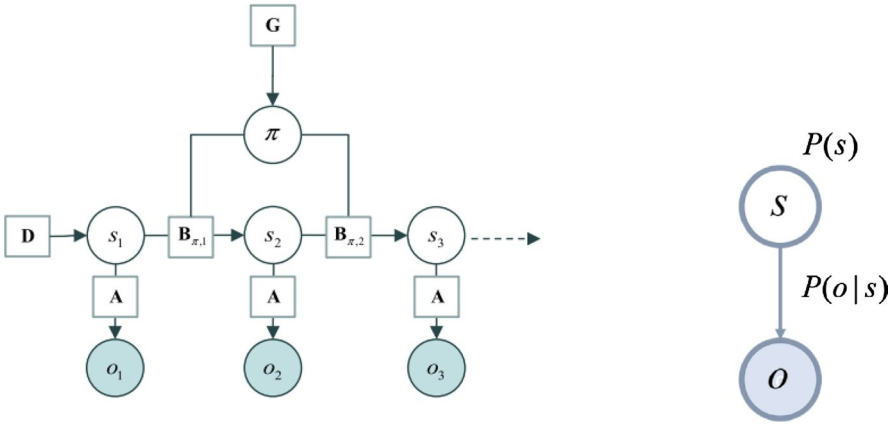

Fig. 3 Two generative models. The left-side generative model embeds inference about hidden states given outcomes within a processing hierarchy with temporal depth (from Friston et al. 2017, p. 386). This means that it becomes possible to infer future outcomes based on actions yet to be selected by the agent. The right-side generative model is an illustration of a generative model that cannot function as a search plan, for it cannot perform inferences over how changing states in the future impacts on outcomes given action (from Parr et al. 2018, p. 3)

on inferred action policies (represented by π). This equips *E.coli* with the ability to make inferences about how its environment will impact on its sensory states given the kind of action it undertakes. Hence, under the FEP, swimming straight (rather than tumbling) suggests that *E.coli* have beliefs about expected future states, and how these will be beneficial for adaptive behavior. In other words, $B\pi_1$ functions as prior beliefs about expected sensory outcomes o_2. At the top of this generative model there is a prior belief (G) about policies (π). This means that at the most foundational level, *E.coli* are naturally biased towards inferring actions that will have the effect of minimising expected free energy in the future. This is what it means to have a search plan under the FEP.

Clark (2017) says that it is problematic to talk of models in this case. We think this is correct, but only if the target is the kind of generative model we see in the right-side of Fig. 3. The reason for this is that this generative model is only able to infer hidden states based on sensory outcomes. If such an agent exists, it would not be able to act—the reason being that it is not equipped with beliefs about expected future states. Indeed, such a system would be a system that must entirely and relentlessly infer its best guesses about the world, and update its priors to fit new incoming evidence. Biological systems are however not such systems—they do not only maximise evidence for their posterior distributions of the relation between outcomes and causes. Biological systems are active systems. They actively monitor and react to perturbations that challenge homeostatic variables, which may, from time to time, go out of bounds (Kirchhoff et al. 2018). This provides one reason for why understanding the operations of organisms over time requires appeal to generative models with temporal depth. Hence, *E.coli* are not the kind of organism that can be understood without reference to probabilistic search plans (under the FEP).

It is important for forestall a possible confusion here. It is not uncommon to hear objections to the FEP along roughly these lines: why think that organisms are optimisers rather than satisficers? Why think that organisms are optimal or become close-to-optimal models of their environment at all? That is, how do we decide on how far from optimal a matter of dispute is Bowers and Davis (2012)? Or, how far from or how close to optimal is some sequence of action? Talk of optimality in the FEP should not be interpreted as if organisms are able to produce optimal inferences, where prior probabilities about the world match the actual distribution of some attributes in the environment. Solutions that are *good enough* will still promote adaptive interchanges with the environment, and thus conform with the FEP. Being a good enough model is still to be a good regulator—a good enough regulator. So, even if a system produces good enough solutions to environmental pressures, it will be a system that is able to reduce surprise (on average over time). Contrary to the universal ethology challenge, then, the FEP provides a formal rationale for the claim that inference and probabilistic search plans lie at the explanatory base of adaptive self-organisation.

Conclusion

If the FEP is true, biological and cognitive characteristics can be methodologically understood as involving approximate Bayesian inference and probabilistic search plans (i.e., generative models). It is these constructs, underwriting the minimisation of free energy, that are utilised to establish a unified theoretical framework for the study of biology and cognitive science. We raised a specific challenge to this project; the universal ethology challenge. This challenge states that the FEP cannot unify biology and cognition, for life (or adaptive self-organisation) itself does not require that organisms minimise free energy via approximate Bayesian inference.

Here we have provided a rationale for thinking that it is possible to show that the FEP can overcome this universal ethology challenge. First, we have argued that there is a notion of inference within the FEP that tracks relations of correlation (or covariance) between variables and events. We pointed out that this notion of inference trades in Shannon information, or relative entropy. This is particularly important for it is generally taken to be the case that this notion of inference (and information) does not add new and problematic concepts to the domains of biology and cognitive science. Note also that even if the examples often make it seem as if biological processes must explicitly compute values when engaged in approximate Bayesian inference, the appeal to correlational forms of inference captures an important observation; namely, that approximate Bayesian inference can be performed without the system having to perform explicit computations over probability distributions (e.g., in the topological structure of neurons in V1 tuned to cardinal orientations of objects in natural scenes). Second, the universal ethology challenge also pressures the central idea of adaptive self-organisation unfolding on the basis of search plans. Here we have argued that search plans within the FEP takes on a probabilistic notion, and that evidence of adaptive self-organisation is evidence for the presence of search plans (when unpacked in a specific way as generative models equipped

with priors over policies). We think our arguments provide reason to believe that the FEP is not threatened by the universal ethology challenge. This we conclude brings the FEP a step closer to realising its ambitions of unifying biology and cognitive science.

Acknowledgement Kirchhoff's work was supported by an Australian Research Council Discovery Project "Mind in Skilled Performance" (DP170102987). Van Es's work was supported by the Research Foundation Flanders (Grant No. 1124818N). We wish to thank Stephen Mann, Mads Julian Dengsø, and Ross Pain for comments on previous drafts of this manuscript. Finally, we would like to thank two reviewers for providing us with valuable insights that have greatly improved the quality of the paper.

References

Auletta G (2013) Information and metabolism in bacterial chemotaxis. Entropy. https://doi.org/10.3390/e15010311

Bowers J, Davies C (2012) Bayesian just-so stories in psychology and neuroscience. Psychol Bull 138(3):389–414. https://doi.org/10.1037/a0026450

Calvo P, Friston K (2017) Predicting green: really radical (plant) predictive processing. J R Soc Interface 14(131):20170096. https://doi.org/10.1098/rsif.2017.0096

Clark A (2017) How to knit your own Markov blanket: resisting the second law with metamorphic minds. In: T Metzinger, W Wiese (eds) Philosophy and predictive processing: 3. Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958573031

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–253

Colombo M, Elkin L, Hartmann S (2018) Being realist about bayes, and the predictive processing theory of mind. Br J Philos Sci. https://doi.org/10.1093/bjps/axy059

Conant R, Ashby R (1970) Every good regulator of a system must be a good model of that system. Int J Syst Sci 1:89–97

Constant A, Clark A, Kirchhoff MD, Friston K (2020) Extended active inference: constructing predictive cognition beyond skulls. Mind Lang. https://doi.org/10.1111/mila.12330

Corcoran AW, Pezzula G, Hohwy J (2020) From allostatic agents to counterfactual cognisers: active inference. Biol Regul Origins Cognit 35(32):1–45

Dayan P, Hinton G, Neal R (1994) The Helmholtz machine. Neural Comput 7:889–904

Engel A, Fries P, Singer W (2001) Dynamics predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci 2:704–716

Friston K, Da Costa L, Hafner D, Hesp C, Parr T (2020) Sophisticated inference. arXiv:2006.04120

Friston K (2019) A free energy principle for a particular physics. Unpublished manuscript.

Friston KJ, Rosch R, Parr T, Price C, Bowman H (2017) Deep temporal models and active inference. Neurosci Biobehav Rev 77:388–402

Friston K, Rigoli F, Ognibene D, Mathys C, Fitzgerald T, Pezzulo G (2015) Active inference and epistemic value. Cognit Neurosci. https://doi.org/10.1080/17588928.2015.1020053

Friston K (2013) Life as we know it. J R Soc Interface. https://doi.org/10.1098/rsif.2013.0475

Friston K, Thornton C, Clark A (2012) Free-energy minimization and the dark-room problem. Front Psychol 3(130):1–7

Friston K (2011) Embodied inference: Or 'I Think Therefore I Am, If I Am What I Think.' In: Tschacher W, Bergomi C (eds) The implications of embodiment (cognition and communication). Imprint Academic, Exeter, pp 89–125

Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127–138

Friston K (2009) The free-energy principle: a rough guide to the brain? Trends Cognit Sci. https://doi.org/10.1016/j.tics.2009.04.005

Friston K, Stephan KE (2007) Free energy and the brain. Synthese 159:417–458

Frith C, Friston K (2015) A duet for one. Conscious Cognit 36:390–405

Girshick A, Landy M, Simoncelli E (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nat Neurosci 14:926–932
Godfrey-Smith P (2016) Other minds: the octopus, the sea, and the deep origins of consciousness. Farrar, Straus and Giroux, New York
Godfrey-Smith P (2007) Information in biology. The Cambridge companion to the philosophy of biology. Cambridge University Press, Cambridge
Godfrey-Smith P (1996) Complexity and the function of mind in nature. Cambridge University Press, Cambridge
Haken H (1983) Synergetics: an introduction. Nonequilibrium phase transition and self-organisation in physics, chemistry and biology. Springer, Berlin
Hesp C, Ramstead M, Constant A, Badcock P, Kirchhoff MD, Friston K (2019) A multi-scale view of the emergent complexity of life: a free energy proposal. In: Price M et al (eds) Evolution, development, and complexity: multiscale models in complex adaptive systems. Springer, Berlin
Hohwy J (2016) The self-evidencing brain. Nous 50(2):259–285
Hohwy J (2013) The predictive mind. Oxford University Press, Oxford
Kirchhoff MD, Kiverstein J (2020) Attuning to the world: the diachronic constitution of the extended conscious mind. Front Psychol. https://doi.org/10.3389/fpsyg.2020.01966
Kirchhoff MD, Kiverstein J (2019a) Extended consciousness and predictive processing: a third-wave view. Routledge, London
Kirchhoff MD, Kiverstein J (2019b) How to demarcate the boundaries of mind: a Markov blanket proposal. Synthese. https://doi.org/10.1007/s11229-019-02370-y
Kirchhoff MD, Robertson I (2018) Enactivism and predictive processing: a non-representational view. Philos Explor. https://doi.org/10.1080/13869795.2018.1477983
Kirchhoff M, Parr T, Palacios E, Friston K, Kiverstein J (2018) The Markov blankets of life: autonomy, active inference and the free energy principle. J R Soc Interface. https://doi.org/10.1098/rsif.2017.0792
Kirchhoff MD (2018) Hierarchical Markov blankets and adaptive active inference. Phys Life Rev. https://doi.org/10.1016/j.plrev.2017.09.001
Kirchhoff MD, Froese T (2017) Where there is life there is mind: in support of a strong life-mind continuity thesis. Entropy. https://doi.org/10.3390/e19040169
Kirchhoff MD (2015) Extended cognition & the causal-constitutive fallacy. Philos Phenomenol Res 90(2):320–360
Knill DC, Pouget A (2004) The Bayesian Brain: the role of uncertainty in neural coding and computation. Trends in Neurosci 27(12):712–719
Laland K, Matthews B, Feldman W (2016) An introduction to niche construction theory. Evol Ecol 30:191–202
Levins R (1966) The strategy of model building in population biology. In: Sober E (ed) Conceptual issues in evolutionary biology. The MIT Press, Cambridge, pp 18–27
Linson A, Clark A, Ramamoorthy S, Friston K (2018) The active inference approach to ecological perception: general information dynamics for natural and artificial embodied cognition. Front Robot AI. https://doi.org/10.3389/frobt.2018.00021
Lyon P, Keijzer F (2007) The human stain: Why cognitivism can't tell us what cognition is and what it does. In: Wallace B (ed) The mind, the world and the body. Imprint Academics, Exeter, pp 132–156
Millidge B, Tschantz A, Buckley C (2020) Whence the expected free energy? https://arxiv.org/pdf/2004.08128.pdf
Odling-Smee J, Laland K, Feldman M (2003) Niche construction: the neglected process in evolution. Princeton University Press, Princeton
Orlandi N (2017) The innocent eye: why vision is not a cognitive process. Oxford University Press, Oxford
Palacios ER, Razi A, Parr T, Kirchhoff M, Friston K (2020) On Markov blankets and hierarchical self-organisation. J Theor Biol 486:110089. https://doi.org/10.1016/j.jtbi.2019.110089
Parr T, Friston K (2019) Generalised free energy and active inference. Biol Cybern 113:495–513
Parr T, Rees G, Friston KJ (2018) Computational neuropsychology and Bayesian inference. Front Hum Neurosci. https://doi.org/10.3389/fnhum.2018.00061
Parr T, Friston K (2017) Working memory, attention, and salience in active inference. Sci Rep 7:14678. https://doi.org/10.1038/s41598-017-15249-0
Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical visual-field effects. Nat Neurosci 2:79–87

Ramstead MJ, Friston KJ, Hipólito I (2020) Is the free-energy principle a formal theory of semantics? From variational density dynamics to neural and phenotypic representations. arXiv:2007.09291
Ramstead M, Kirchhoff MD, Constant A, Friston K (2019) Multiscale integration: beyond internalism and externalism. Synthese. https://doi.org/10.1007/s11229-019-02115-x
Ramstead M, Badcock P, Friston KJ (2017) Answering Schrödinger's question: a free-energy formulation. Phys Life Rev. https://doi.org/10.1016/j.plrev.2017.09.001
Schrödinger E (1944) What is life? Cambridge University Press, Cambridge
Seth AK (2015) The cybernetic brain: from interoceptive inference to sensorimotor contingencies. In: Metzinger T, Windt JM (eds) Open MIND: 35(T). Frankfurt am Main, Germany: MIND Group
Sterelny K (2005) Made by each other: Organisms and their environment. Biol Philos 20:21–36
Teufel C, Fletcher P (2020) Forms of predictions in the nervous system. Nature 21:231–242
Weisberg M (2006) Forty years of 'the strategy': levins on model building and idealization. Biol Philos 21:623–645
Wiese W (2017) What are the contents of representations in predictive processing? Phenomenol Cognit Sci 16(4):715–736
Williams D (2018) Predictive coding and thought. Synthese 197:1749–1775