

Whence the Expected Free Energy?

Beren Millidge

beren@millidge.name

School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, U.K.

Alexander Tschantz

tschantz.alec@gmail.com

Sackler Center for Consciousness Science, School of Engineering and Informatics, University of Sussex, Falmer, Brighton, BN1 9RH, U.K.

Christopher L. Buckley

C.L.Buckley@sussex.ac.uk

Evolutionary and Adaptive Systems Research Group, School of Engineering and Informatics, University of Sussex, Falmer, Brighton, BN1 9RH, U.K.

The expected free energy (EFE) is a central quantity in the theory of active inference. It is the quantity that all active inference agents are mandated to minimize through action, and its decomposition into extrinsic and intrinsic value terms is key to the balance of exploration and exploitation that active inference agents evince. Despite its importance, the mathematical origins of this quantity and its relation to the variational free energy (VFE) remain unclear. In this letter, we investigate the origins of the EFE in detail and show that it is not simply “the free energy in the future.” We present a functional that we argue is the natural extension of the VFE but actively discourages exploratory behavior, thus demonstrating that exploration does not directly follow from free energy minimization into the future. We then develop a novel objective, the free energy of the expected future (FEEF), which possesses both the epistemic component of the EFE and an intuitive mathematical grounding as the divergence between predicted and desired futures.

1 Introduction

The free-energy principle (FEP) (Friston, 2010; Friston & Ao, 2012; Friston, Kilner, & Harrison, 2006) is an emerging theory from theoretical neuroscience that offers a unifying explanation of the dynamics of self-organizing systems (Friston, 2019; Parr, Da Costa, & Friston, 2020). It proposes that such systems can be interpreted as embodying a process of variational inference that minimizes a single information-theoretic objective: the variational free-energy (VFE). In theoretical neuroscience, the FEP translates into an

elegant account of brain function (Friston, 2003, 2005, 2008a, 2008b; Friston, Trujillo-Barreto, & Daunizeau, 2008), extending the Bayesian brain hypothesis (Deneve, 2005; Doya, Ishii, Pouget, & Rao, 2007; Knill & Pouget, 2004) by postulating that the neural dynamics of the brain perform variational inference. Under certain assumptions about the forms of the densities embodied by the agent, this theory can even be translated down to the level of neural circuits in the form of a biologically plausible neuronal process theory (Bastos et al., 2012; Friston, 2008a; Kanai, Komura, Shipp, & Friston, 2015; Shipp, 2016; Spratling, 2008).

Action is then subsumed into this formulation, under the name of *active inference* (Friston, 2011; Friston & Ao, 2012; Friston, Daunizeau, & Kiebel, 2009) by mandating that agents act so as to minimize the VFE with respect to action (Buckley, Kim, McGregor, & Seth, 2017; Friston et al., 2006). This casts action and perception as two aspects of the same imperative of free-energy minimization, resulting in a theoretical framework for control that applies to a variety of continuous-time tasks (Baltieri & Buckley, 2017, 2018; Calvo & Friston, 2017; Friston, Mattout, & Kilner, 2011; Millidge, 2019b).

Recent work has extended these ideas to account for inference over temporally extended action sequences. (Friston & Ao, 2012; Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2017; Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016; Friston et al., 2015; Tschantz, Seth, & Buckley, 2019). Here it is assumed that rather than action minimizing the instantaneous VFE, sequences of actions (or policies) minimize the cumulative sum over time of a quantity called the *expected free energy* (EFE) (Friston et al., 2015). Active inference using the EFE has been applied to a wide variety of tasks and applications, from modeling human and animal choice behavior (FitzGerald, Schwartenbeck, Moutoussis, Dolan, & Friston, 2015; Friston et al., 2015; Pezzulo, Cartoni, Rigoli, Pio-Lopez, & Friston, 2016), simulating visual saccades and other “epistemic foraging behavior” (Friston, Lin, et al., 2017; Friston, Rosch, Parr, Price, & Bowman, 2018; Mirza, Adams, Mathys, & Friston, 2016; Parr & Friston, 2017a, 2018a), solving reinforcement learning benchmarks (Çatal, Verbelen, Nauta, De Boom, & Dhoedt, 2020; Millidge, 2019a, 2020; Tschantz, Baltieri, Seth, & Buckley, 2019; Ueltzhöffer, 2018; van de Laar & de Vries, 2019), to modeling psychiatric disorders as cases of aberrant inference (Cullen, Davey, Friston, & Moran, 2018; Mirza, Adams, Parr, & Friston, 2019; Parr & Friston, 2018b). Like the continuous-time formulation, active inference also comes equipped with a biologically plausible process theory with variational update equations, which have been argued to be homologous with observed neural firing patterns (Friston, FitzGerald, et al., 2017; Friston, Parr, & de Vries, 2017; Parr, Markovic, Kiebel, & Friston, 2019).

A key property of the EFE is that it decomposes into both an extrinsic, value-seeking and an intrinsic (epistemic), information-seeking term (Friston et al., 2015). The latter mandates active inference agents to resolve uncertainty by encouraging the exploration of unknown regions of the environment, a property that has been extensively investigated (Friston,

FitzGerald, et al., 2017a; Friston et al., 2015; Schwartenbeck, FitzGerald, Dolan, & Friston, 2013; Schwartenbeck et al., 2019). The fact that intrinsic drives naturally emerge from this formulation is argued as an advantage over other formulations that typically encourage exploration by adding ad hoc exploratory terms to their loss function (Burda et al., 2018; Mohamed & Rezende, 2015; Oudeyer & Kaplan, 2009; Pathak, Agrawal, Efros, & Darrell, 2017). While the EFE is often described as a straightforward extension to the free energy principle that can account for prospective policies and is typically expressed in similar mathematical form (Da Costa et al., 2020; Friston, FitzGerald, et al., 2017; Friston et al., 2015; Parr & Friston, 2017b, 2019), its origin remains obscure. Minimization of the EFE is sometimes motivated by a *reductio ad absurdum* argument following from the VFE (Friston et al., 2015; Parr & Friston, 2019) in that agents are driven to minimize the VFE, and therefore the only way they can act is to minimize their free energy into the future. Since the future is uncertain, however, instead they must minimize the expected free energy. Central to this logic is the formal identification of the VFE with the EFE.

In this letter, we set out to investigate the origin of the EFE and its relations with the VFE. We provide a broader perspective on this question, showing that the EFE is not the only way to extend the VFE to account for action-conditioned futures. We derive an objective that we believe to be a more natural analog of the VFE, which we call the *free energy of the future* (FEF), and make a detailed side-by-side comparison of the two functionals. Crucially, we show that the FEF actively discourages information-seeking behavior, thus demonstrating that epistemic terms do not necessarily arise simply from extending the VFE into the future. We then investigate the origin of the epistemic term of the EFE and show that the EFE is just the FEF minus the negative of the epistemic term in the EFE, which thus provides a straightforward perspective on the relation between the two functionals. We propose our own mathematically principled starting point for action selection under active inference: the divergence between desired and expected futures, from which we obtain a novel functional, the free-energy of the expected future (FEEF), which has close relations to the generalized free energy (Parr & Friston, 2019). This functional has a natural interpretation in terms of the divergence between a veridical and a biased generative model; it allows use of the same functional for both inference and policy selection, and it naturally decomposes into an extrinsic value term and an epistemic action term, thus maintaining the attractive exploratory properties of EFE-based active inference while also possessing a mathematically principled starting point with an intuitive interpretation.

2 The Variational Free Energy

The variational free energy (VFE) is a core quantity in variational inference and constitutes a tractable bound on both the log model evidence and the Kullback-Leibler (KL) divergence between prior and posterior (Beal, 1998;

Blei, Kucukelbir, & McAuliffe, 2017; Fox & Roberts, 2012; Wainwright & Jordan, 2008). (For an in-depth motivation of the VFE and its use in variational inference, see appendix A.)

The VFE, defined at time t , denoted by F_t , is given by,

$$\begin{aligned} F_t &= \mathbf{D}_{KL}[Q(x_t|o_t; \phi) || p(o_t, x_t)] \\ &= \mathbb{E}_{Q(x_t|o_t; \phi)} \left[\ln \frac{Q(x_t|o_t; \phi)}{p(o_t, x_t)} \right]. \end{aligned} \quad (2.1)$$

The agent receives observations o_t and must infer the values of hidden states x_t . The agent assumes that the environment evolves according to a Markov process so that the distribution over states at the current time step only depends on the state at the previous time step, and that the observation generated at the current time step depends only on the state at the current time step. Given a distribution over a trajectory of states and observations and under Markov assumptions, it can be factorized as follows: $p(o_{0:T}, x_{0:T}) = p(s_0) \prod_{t=0}^T p(o_t|s_t)p(s_{t+1}|s_t)$. In this letter, we also consider inference over future states and observations that have yet to be observed. Such future variables are denoted o_τ or x_τ where $\tau > t$. To avoid dealing with infinite sums, agents only consider futures up to some finite time horizon, denoted T . $Q(x_t|o_t; \phi)$ denotes an approximate posterior density parameterized by ϕ , which, during the course of variational inference, is fit as closely as possible to the true posterior. Note that there is a slight difference in notation here compared to that usually used in variational inference. Normally the approximate posterior is written as $Q(x_t; \phi)$ without the dependence on o made explicit. This is because the variational posterior is not a direct function of observations, but rather the result of an optimization process that depends on the observations. Here, we make the dependence on o explicit to keep a clear distinction between the variational posterior $Q(x_t|o_t; \phi)$, obtained through optimization of the variational parameters ϕ , and the variational prior $Q(x_t) = \mathbb{E}_{p(s_t|s_{t-1})}[Q(s_{t-1}|o_{t-1}; \phi)]$, obtained by mapping the previous posterior through the transition dynamics. Throughout this letter, we assume that inference is occurring in a discrete-time partially observed Markov decision process (POMDP). This is to ensure compatibility with the EFE formulation later, which is also situated within discrete-time POMDPs.¹

¹It is important to note that the original FEP was formulated in continuous time with generalized coordinates (Friston, 2008a; Friston et al., 2006) (where the hidden states are augmented with their temporal derivatives up to a theoretically infinite order). The generalized coordinates mean that the agent is effectively performing variational inference over a Taylor-expanded future trajectory instead of a temporally instant hidden state (Friston, 2008a; Friston et al., 2008). Action is derived by minimizing the gradients of the instantaneous VFE with respect to action, which requires the use of a forward model. More recent work on active inference and the FEP returns to the continuous-time formulation (Friston,

The utility of the VFE for inference comes from the fact that the VFE is equal to the divergence between true and approximate posteriors up to a constant: $\mathbf{F}_t \geq \mathbf{D}_{KL}[Q(x_t|o_t; \phi) || p(x_t|o_t)]$. Thus, minimizing \mathbf{F}_t with respect to the parameters of the variational distribution makes $Q(x_t; \phi)$ a good approximation of the true posterior.

One can also motivate the VFE as a technique to estimate model evidence. Log model evidence is a key quantity in Bayesian inference but is often intractable, meaning it cannot be computed directly. Intuitively, the log model evidence scores the likelihood of the data under a model and thus provides a direct measure of the quality of a model. Under the free energy principle, minimizing the negative log model evidence (or surprisal) is the ultimate goal of self-organizing systems (Friston & Ao, 2012; Friston et al., 2006). The VFE provides an upper bound on the log model evidence. This can be shown by importance-sampling the model evidence with respect to the approximate posterior and applying Jensen’s inequality:

$$\begin{aligned}
 -\ln p(o_t) &= -\ln \int dx_t p(o_t, x_t) \\
 &= -\ln \int dx_t p(o_t, x_t) \frac{Q(x_t|o_t; \phi)}{Q(x_t|o_t; \phi)} \\
 &\leq -\int dx_t Q(x_t|o_t; \phi) \ln \frac{p(o_t, x_t)}{Q(x_t|o_t; \phi)} \\
 &\leq \mathbf{D}_{KL}[Q(x_t|o_t; \phi) || p(o_t, x_t)] \\
 &\leq \mathbf{F}_t.
 \end{aligned}$$

Since the VFE is an upper bound on the log model evidence (or surprisal), as the VFE is minimized, it becomes an increasingly accurate estimate of the surprisal. To get a sense of the properties of the VFE, we showcase the following decomposition:

$$\begin{aligned}
 \mathbb{F} &= \mathbf{D}_{KL}[Q(x_t|o_t; \phi) || p(o_t, x_t)] \\
 &= \mathbb{E}_{Q(x_t|o_t; \phi)} \left[\ln \frac{Q(x_t|o_t; \phi)}{p(o_t, x_t)} \right] \\
 &= \underbrace{-\mathbb{E}_{Q(x_t|o_t; \phi)} [\ln p(o_t|x_t)]}_{\text{Accuracy}} + \underbrace{\mathbf{D}_{KL}[Q(x_t|o_t; \phi) || p(x_t)]}_{\text{Complexity}}. \tag{2.2}
 \end{aligned}$$

This decomposition is the one typically used to compute the VFE in practice and has a straightforward interpretation. Specifically, minimizing the

2019; Parr, Da Costa, & Friston, 2020) and the conclusions drawn in this article may look different in the continuous-time domain.

negative accuracy (and thus maximizing accuracy) ensures that the observations are as likely as possible under the states, x_t , predicted by the variational posterior while simultaneously minimizing the complexity term, which is a KL-divergence between the variational posterior and the prior. Thus, the goal is to keep the posterior as close to the prior as possible while maximizing accuracy. Effectively, the complexity term acts as an implicit regularizer, reducing the risk of overfitting to any specific observation.

3 The Expected Free Energy

While variational inference as presented only allows us to perform inference at the current time given observations, it is possible to extend the formalism to allow for inference over actions or policies in the future.

To achieve this extension, a variational objective is required that can be minimized contingent on future states and policies, which will allow the problem of adaptive action selection to be reformulated as a process of variational inference. To do this, the formalism must be extended in two ways. First, the generative model is augmented to include actions a_τ , and policies, which are sequences of actions $\pi = [a_1, a_2 \dots a_T]$. The action taken at the current time can affect future states, and thus future observations. In order to transform action selection into an inference problem, policies are treated as an inferred distribution $Q(\pi)$ that is optimized to meet the agent's goals. The second extension required is to translate the notion of an agent's goals into this probabilistic framework. Active inference encodes an agent's goals as a desired distribution over observations $\tilde{p}(o_{\tau:T})$. We denote the biased distribution using a tilde over the probability density \tilde{p} rather than the random variable to make clear that the random variables themselves are unchanged; it is only the agent's subjective distribution over the variables that is biased.² This distribution is then incorporated into a biased generative model of the world $\tilde{p}(o_\tau, x_\tau) \approx \tilde{p}(o_\tau)Q(x_\tau|o_\tau)$,³ where we have additionally made the assumption that the true posterior can be well approximated with the variational posterior: $p(x_\tau|o_\tau) \approx Q(x_\tau|o_\tau)$ which simply states that the variational inference procedure was successful.⁴ Active

² It is important to note that this encoding of preferences through a biased generative model is unique to active inference. Other variational control schemes (Levine, 2018; Rawlik, Toussaint, & Vijayakumar, 2013; Rawlik, 2013; Theodorou, Buchli, & Schaal, 2010; Theodorou & Todorov, 2012) instead encode desires through binary optimality variables and optimize the posterior given that the optimal path was taken. The relation between these frameworks is explored further in Millidge, Tschantz, Seth, and Buckley (2020).

³ Some more recent work (Da Costa et al., 2020; Friston, 2019) prefers an alternative factorization of the biased generative model in terms of an unbiased likelihood and a biased prior state distribution $\tilde{p}(o_\tau, x_\tau) = p(o_\tau|x_\tau)\tilde{p}(x_\tau)$. This leads to a different decomposition of the EFE in terms of risk and ambiguity (see appendix B) but which is mathematically equivalent to the factorization described here.

⁴ For additional information on the effect of this assumption, see appendix D.

inference proceeds by inferring a variational policy distribution $Q(\pi)$ that maximizes the evidence for this biased generative model. Intuitively, this approach turns the action selection problem on its head. Instead of saying, “I have some goal; what do I have to do to achieve it?” the active inference agent asks: “Given that my goals were achieved, what would have been the most probable actions that I took?”

A further complication of extending VFE into the future comes from future observations. While agents have access to current observations (or data) for planning problems, they must also reason about unknown future observations. This is dealt with by taking the expectation of the objective with respect to predicted observations o_τ drawn from the generative model.

In the active inference framework, the goal is to infer a variational distribution over both hidden states and policies that maximally fit to a biased generative model of the future. The framework defines the variational objective function to be minimized, the expected free energy, from time τ until the time horizon T , which is denoted \mathcal{G} :

$$\mathcal{G} = \mathbb{E}_{Q(o_{\tau:T}, x_{\tau:T}, \pi)} [\ln Q(x_{\tau:T}, \pi) - \ln \tilde{p}(o_{\tau:T}, x_{\tau:T})].$$

A temporal mean-field factorization of the approximate posterior and of the generative model is assumed such that $Q(x_{\tau:T}, \pi) \approx Q(\pi) \prod_{\tau} Q(x|\pi)$ and $\tilde{p}(o_{\tau:T}, x_{\tau:T}) \approx \prod_{\tau} \tilde{p}(o_\tau)Q(x_\tau|o_\tau)$. This factorization neatly severs the temporal dependencies between time steps. Given these assumptions, inferring the optimal $Q(\pi)$, turns out to be relatively straightforward:

$$\begin{aligned} \mathcal{G} &= \mathbb{E}_{Q(o_{\tau:T}, x_{\tau:T}, \pi)} [\ln Q(x_{\tau:T}, \pi) - \ln \tilde{p}(o_{\tau:T}, x_{\tau:T})] \\ &= \mathbb{E}_{Q(o_{\tau:T}, x_{\tau:T}|\pi)Q(\pi)} [\ln Q(x_{\tau:T}|\pi) + \ln Q(\pi) - \ln \tilde{p}(o_{\tau:T}, x_{\tau:T})] \\ &= \mathbb{E}_{Q(\pi)} [\ln Q(\pi) - \mathbb{E}_{Q(o_{\tau:T}, x_{\tau:T}|\pi)} \left[\sum_{\tau} [\ln Q(x|\pi) - \ln \tilde{p}(o_\tau, x_\tau)] \right]] \\ &= \mathbf{D}_{KL} [Q(\pi) \| e^{-\sum_{\tau} \mathcal{G}_\tau(\pi)}], \end{aligned}$$

where $\mathcal{G}_\tau(\pi) = \mathbb{E}_{Q(o_\tau, x_\tau|\pi)} [\ln Q(x_\tau|\pi) - \ln \tilde{p}(o_\tau, x_\tau)]$ is defined to be the EFE for a single time step τ . From the KL-divergence above, it follows that the optimal variational policy distribution $Q^*(\pi)$ is simply the path integral into the future of the expected free energies for each individual time step,

$$Q^*(\pi) = \sigma \left(\sum_t^T \mathcal{G}_t(\pi) \right),$$

where $\sigma(x)$ is a softmax function. This implies that to infer the optimal policy distribution, it suffices to minimize the sum of expected free energies for

each time step into the future. Inference proceeds by using the generative model to roll out predicted futures, computing the EFE of those futures, and then selecting policies that minimize the sum of the expected free energies. Since under temporal mean field assumptions, trajectories decompose into a sum of time steps, it is sufficient for the rest of the letter to only consider a single time step τ .

To gain an intuition for the EFE, we showcase the following decomposition:

$$\begin{aligned} \mathcal{G}_\tau(\pi) &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau, x_\tau)] \\ &\approx \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau) - \ln Q(x_\tau | o_\tau)] \\ &\approx \underbrace{-\mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln \tilde{p}(o_\tau)]}_{\text{Extrinsic Value}} - \underbrace{\mathbb{E}_{Q(o_\tau)} \mathbf{D}_{KL}[Q(x_\tau | o_\tau) || Q(x_\tau | \pi)]}_{\text{Epistemic Value}}. \quad (3.1) \end{aligned}$$

While the EFE admits many decompositions (see appendix B for a comprehensive overview), the one presented in equation 3.1 is perhaps the most important because it separates the EFE into an extrinsic, goal-directed term (sometimes also called instrumental value in the literature) and an intrinsic, information-seeking term.⁵ The first term requires agents to maximize the likelihood of the desired observations $\tilde{p}(o_\tau)$ under beliefs about the future. It thus directs an agent to act to maximize the probability of its desires occurring in the future. It is called the extrinsic value term since it is the term in the EFE that accounts for the agent's preferences.

The second term in equation 3.1 is the expected information gain, which is often termed the epistemic value since it quantifies the amount of information gained by visiting a specific state. Since the information gain is negative, minimizing the EFE as a whole mandates maximizing the expected information gain. This drives the agent to maximize the divergence between its posterior and prior beliefs, thus inducing the agent to take actions that maximally inform their beliefs and reduce uncertainty. It is the combination of extrinsic and intrinsic value terms that belies active inference's claim to have a principled approach to the exploration-exploitation dilemma (Friston, FitzGerald, et al., 2017; Friston et al., 2015).

The idea of maximizing expected information gain or "Bayesian surprise" (Itti & Baldi, 2009) to drive exploratory behavior has been argued for in neuroscience (Baldi & Itti, 2010; Ostwald et al., 2012) and has been regularly proposed in reinforcement learning (Houthoofd et al., 2016; Still & Precup, 2012; Sun, Gomez, & Schmidhuber, 2011; Tschantz, Millidge, Seth,

⁵The approximation in the final line of equation 3.1 is that we assume that the true and approximate posteriors are the same $Q(x_\tau | o_\tau) \approx p(x_\tau | o_\tau)$. Without this assumption, you obtain an additional KL-divergence between the true and approximate posterior, which exactly quantifies the discrepancy between them (see appendices B and D for more detail).

& Buckley, 2020). It is important to note, however, that in these prior works, information gain has often been proposed as an ad hoc addition to an existing objective function with only the intuitive justification of boosting exploration. In contrast, expected information gain falls naturally out of the EFE formalism, arguably lending the formalism a degree of theoretical elegance.

4 Origins of the EFE

Given the centrality of the EFE to the active inference framework, it is important to explore the origin and nature of this quantity. The EFE is typically motivated through a *reductio ad absurdum* argument (Friston et al., 2015; Parr & Friston, 2019).⁶ The logic is as follows. Agents have prior beliefs over policies that drive action selection. By the FEP, all states of an organism, including those determining policies, must change so as to minimize free energy. Thus, the only self-consistent prior belief over policies is that the agent will minimize free energy into the future through its policy selection process. If the agent did not have such a prior belief, then it would select policies that did not minimize the free energy into the future and would thus not be a free energy minimizing agent. This logic requires a well-defined notion of the free energy of future states and observations given a specific policy. The active inference literature implicitly assumes that the EFE is the natural functional that fits this notion (Friston, FitzGerald, et al., 2017; Friston et al., 2015). In the following section, we argue that the EFE is not in fact the only functional that can quantify the notion of the free energy of policy-conditioned futures, and indeed we propose a different functional, the free energy of the future, which we argue is a more natural extension of the VFE to account for future states.

4.1 The Free Energy of the Future. We argue that the natural extension of the free energy into the future must possess direct analogs to the two crucial properties of the VFE: it must be expressible as a KL-divergence between a posterior and a generative model, such that minimizing it causes the variational density to better approximate the true posterior, and it must also bound the log model evidence of future observations. Bounding the log model evidence (or surprisal) is vital since the surprisal is the core quantity that, under the FEP, all systems are driven to minimize. If the VFE extended into the future failed to bound the surprisal, then minimizing this extension would not necessarily minimize surprisal, and thus any agent that minimized such an extension would be in violation of the FEP. Here, we present

⁶An alternative motivation exists that situates the expected free energy in terms of a nonequilibrium steady-state distribution (Da Costa et al., 2020; Friston, 2019; Parr, 2019). This argument reframes everything in terms of a Gibbs free energy, from which the EFE can be derived as a special case. The problem becomes, then, one of the motivation of the Gibbs free energy as an objective function.

a functional that we claim satisfies these desiderata: the free energy of the future (FEF).

We wish to derive an expression for variational free energy at some future time τ that is conditioned on some policy π . In other words, we wish to quantify the free energy that will occur at some future time point, given some sequence of actions. Here, we derive a form of the variational free energy of the future, denoted $\text{FEF}_\tau(\pi)$, by keeping the same terms as the VFE (see equation 2.1), but conditioning the variational distributions on our policy of interest and rewriting for the future time point τ . Additionally, since observations in the future are unknown, we must evaluate our free energy under the expectation of our beliefs about future observations, as in the EFE. We thus define

$$\text{FEF}_\tau(\pi) = \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | o_\tau) - \ln \tilde{p}(o_\tau, x_\tau)].$$

Since this equation is simply the KL-divergence between the variational posterior and the generative model, it satisfies the first desideratum. We next investigate the properties of the FEF by showcasing one key decomposition. As with the the VFE, we can then split the FEF into an energy and an entropy or an accuracy and complexity term, which corresponds to the extrinsic and epistemic action terms in the EFE:

$$\begin{aligned} \text{FEF}_\tau(\pi) &= \mathbb{E}_{Q(o_\tau | \pi)} \mathbf{D}_{\text{KL}}[Q(x_\tau | o_\tau) || \tilde{p}(o_\tau, x_\tau)] \\ &\approx \underbrace{-\mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln \tilde{p}(o_\tau | x_\tau)]}_{\text{Accuracy}} + \underbrace{\mathbb{E}_{Q(o_\tau | \pi)} \mathbf{D}_{\text{KL}}[Q(x_\tau | o_\tau) || Q(x_\tau | \pi)]}_{\text{Complexity}}. \end{aligned}$$

Unlike the EFE, however, the expected information gain (complexity) term is positive, while in the EFE term, it is negative. Since the objective function, whether EFE or FEF, is to be minimized, we see that using the FEF mandates us to minimize the information gain, while the EFE requires us to maximize it (or minimize the negative information gain). An FEF agent thus tries to maximize its reward while trying to explore as little as possible. While this sounds surprising, it is in fact directly analogous to the complexity term in the VFE, which mandates maximizing the likelihood of an observation, while also keeping the posterior as close as possible to the prior.⁷

4.2 Bounds on the Expected Model Evidence. We next show how the FEF can be derived as a bound on the expected model evidence satisfying the second desideratum. We define the expected model evidence to be a straightforward extension of the model evidence to unknown future states.

⁷An objective functional equivalent to the FEF—the predicted free energy—has also been proposed in Schwöbel, Kiebel, and Marković (2018). See appendix F for more details.

The expected negative log model evidence for a trajectory from the current time step t to some time horizon T is

$$-\mathbb{E}_{Q(o_{t:T}|\pi)}[\ln \tilde{p}(o_{t:T})].$$

This objective states that we wish to maximize the probability (minimize the negative probability) of being in a desired trajectory $\tilde{p}(o_{t:T})$, expected under the distribution of our beliefs about our likely future trajectories $Q(o_{t:T}|\pi)$ under a specific policy π . Given a Markov generative model $p(o_{1:T}, x_{1:T}|\pi) = \prod_t^T p(o_t|x_t)p(x_t|x_{t-1}, \pi)$, and assuming that the approximate posterior factorizes $Q(x_{1:T}|o_{1:T}) = \prod_{t=1}^T Q(x_t|o_t)$, the expected model evidence factorizes across time steps, it suffices to show the derivation for a single time step $\tau > t$ (see appendix C for a full trajectory derivation). We further define $Q(o_\tau, x_\tau|\pi) = Q(o_\tau|\pi)Q(x_\tau|o_\tau) = p(o_\tau|x_\tau)Q(x_\tau|\pi)$. We therefore take the expected model evidence for a single time step and show that the FEF is a bound on this quantity:

$$\begin{aligned} -\mathbb{E}_{Q(o_\tau|\pi)}[\ln \tilde{p}(o_\tau)] &= -\mathbb{E}_{Q(o_\tau|\pi)}\left[\ln \int dx_\tau \tilde{p}(o_\tau, x_\tau)\right] & (4.1) \\ &= -\mathbb{E}_{Q(o_\tau|\pi)}\left[\ln \int dx_\tau \tilde{p}(o_\tau, x_\tau) \frac{Q(x_\tau|o_\tau)}{Q(x_\tau|o_\tau)}\right] \\ &\leq -\mathbb{E}_{Q(o_\tau|\pi)} \int dx_\tau Q(x_\tau|o_\tau) \left[\ln \frac{\tilde{p}(o_\tau, x_\tau)}{Q(x_\tau|o_\tau)}\right] \\ &\leq -\mathbb{E}_{Q(o_\tau, x_\tau|\pi)} \left[\ln \frac{\tilde{p}(o_\tau, x_\tau)}{Q(x_\tau|o_\tau)}\right] \\ &\leq \mathbb{E}_{Q(o_\tau, x_\tau|\pi)} \left[\ln \frac{Q(x_\tau|o_\tau)}{\tilde{p}(o_\tau, x_\tau)}\right] \\ &\leq \mathbb{E}_{Q(o_\tau|\pi)} \mathbf{D}_{KL}[Q(x|o_\tau) \parallel \tilde{p}(o_\tau, x_\tau|\pi)] = \mathbf{FEF}(\pi). \end{aligned}$$

Crucially, this is an upper bound on expected model evidence, which can be tightened by minimizing the FEF. By contrast, returning to the EFE, we see below that since KL-divergences are always ≥ 0 , the expected information gain is always positive, and so the EFE is a lower bound on the expected model evidence:

$$\begin{aligned} \mathcal{G}_\tau(\pi) &= \mathbb{E}_{Q(o_\tau, x_\tau|\pi)}[\ln Q(x_\tau|\pi) - \ln \tilde{p}(o_\tau, x_\tau)] \\ &\approx \underbrace{-\mathbb{E}_{Q(o_\tau, x_\tau|\pi)}[\ln \tilde{p}(o_\tau)]}_{\text{Negative Expected Log Model Evidence}} - \underbrace{\mathbb{E}_{Q(o_\tau|\pi)} \mathbf{D}_{KL}[Q(x_\tau|o_\tau) \parallel Q(x_\tau|\pi)]}_{\text{Expected Information Gain}}. \end{aligned}$$

Since the expected information gain is an expected KL-divergence, it must be ≥ 0 , and thus the negative expected information gain must be ≤ 0 .

Since the EFE aims to minimize negative information gain (thus maximizing positive information gain), we can see that minimizing the EFE actually drives it further from the expected model evidence.⁸

We further investigate the EFE and its properties as a bound in appendix D. Additionally, in appendix E we review other attempts in the literature to derive the EFE as a bound on the expected model evidence and discuss their shortcomings.

4.3 The EFE and the FEF. To get a stronger intuition for the subtle differences between the EFE and the FEF, we present a detailed side-by-side comparison of the two functionals:

$$\text{FEF} = \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | o_\tau) - \ln \tilde{p}(o_\tau, x_\tau)],$$

$$\text{EFE} = \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau, x_\tau)].$$

While the two formulations might initially look very similar, the key difference is the variational term. The FEF, analogous to the VFE, measures the difference between a variational posterior $Q(x_\tau | o_\tau)$ and the generative model $Q(x_\tau | \pi)$. The EFE, on the other hand, measures the difference between a variational prior and the generative model. It is this difference that makes the EFE not a straightforward extension to the VFE for future time steps, and underwrites its unique epistemic value term.

We now demonstrate that both the EFE and the FEF can be decomposed into an expected likelihood, associated with extrinsic value, and an expected KL-divergence between a variational posterior and a variational prior, associated with epistemic value. We factorize the generative model in the FEF into the (biased) likelihood and a variational prior, and factorize the generative model in the EFE into an approximate posterior, and a (biased) marginal:

$$\text{FEF} = \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | o_\tau) - \ln \tilde{p}(o_\tau | x_\tau) - \ln Q(x_\tau | \pi)],$$

$$\text{EFE} = \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau) - \ln Q(x_\tau | o_\tau)].$$

The variational prior and variational posterior can then be combined in both the FEF and the EFE to form epistemic terms. Crucially, the epistemic value term is positive in the FEF and negative in the EFE, meaning that the

⁸There is a slight additional subtlety here involving the fact that there is also a posterior approximation error term that is positive. In general, the EFE functions as an upper bound when the posterior error is greater than the information gain and a lower bound when the posterior error is smaller. Since the goal of variational inference is to minimize posterior error, and EFE agents are driven to maximize expected information gain, we expect this latter condition to occur rarely. For more detail, see appendix D.

FEF penalizes epistemic behavior whereas the EFE promotes it:

$$\begin{aligned}
 \mathbf{FEF} &= \underbrace{-\mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln \tilde{p}(o_\tau | x_\tau)]}_{\text{Extrinsic Value}} + \underbrace{\mathbb{E}_{Q(o_\tau | \pi)} \mathbf{D}_{KL}[Q(x_\tau | o_\tau) || Q(x_\tau | \pi)]}_{\text{Epistemic Value}} \quad (4.2) \\
 \mathbf{EFE} &= \underbrace{-\mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln \tilde{p}(o_\tau)]}_{\text{Extrinsic Value}} - \underbrace{\mathbb{E}_{Q(o_\tau | \pi)} \mathbf{D}_{KL}[Q(x_\tau | o_\tau) || Q(x_\tau | \pi)]}_{\text{Epistemic Value}}.
 \end{aligned}$$

Equation 4.2. demonstrates that the FEF and EFE can be decomposed in similar fashion. We note that the extrinsic value term for the FEF is a likelihood and a marginal for the EFE. The most important difference, however, lies in the sign of the epistemic value term. Since optimizing either the FEF or the EFE requires their minimization, minimizing the FEF mandates us to minimize information gain while the EFE requires us to maximize it. An FEF agent thus tries to maximize its extrinsic value while trying to explore as little as possible. A key question then arises: Where does the negative information gain in the EFE come from?

While this difference in the sign of the expected information gain term may speak to some deep connection between the two quantities, here we offer a pragmatic perspective on the matter. We show that a possible route to the EFE is simply that it is the FEF minus the expected information gain. This implies that the epistemic value term of the EFE arises not from some connection to variational inference but is present by construction:

$$\begin{aligned}
 \mathbf{FEF}_\tau(\pi) - \mathbf{IG}_\tau &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} \left[\ln \left(\frac{Q(x_\tau | o_\tau)}{\tilde{p}(o_\tau, x_\tau)} \right) \right] - \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} \left[\ln \left(\frac{Q(x_\tau | o_\tau)}{Q(x_\tau | \pi)} \right) \right] \\
 &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} \left[\ln \left(\frac{Q(x_\tau | o_\tau) Q(x_\tau | \pi)}{\tilde{p}(o_\tau, x_\tau) Q(x_\tau | o_\tau)} \right) \right] \\
 &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} \left[\ln \left(\frac{Q(x_\tau | \pi)}{\tilde{p}(o_\tau, x_\tau)} \right) \right] \\
 &= \mathbf{EFE}(\pi)_\tau.
 \end{aligned}$$

While this proof illustrates the relation between the EFE and the FEF, it is theoretically unsatisfying as an account of the origin of the EFE. A large part of the appeal of the EFE is that it purports to show that epistemic value arises “naturally” out of minimizing free energy into the future. In contrast, here we have shown that minimizing free energy into the future requires no commitment to exploratory behavior. While this does not question the usefulness of using an information gain term for exploration, or the use of the EFE as a loss function, it does raise questions about the mathematically principled nature of the objective. It is thus not straightforward to see why agents are directly mandated by the FEP to minimize the EFE specifically, as opposed to some other free energy functional. While this fact

may at first appear concerning, we believe it ultimately enhances the power of the formalism by licensing the extension of active inference to encompass other objective functions in a principled manner (Biehl, Guckelsberger, Salge, Smith, & Polani, 2018). In the following section, we propose an alternative objective to the EFE, which results in the same information-seeking epistemic value term, but derives it in a mathematically principled and intuitive way as a bound on the divergence between expected and desired futures.

5 Free Energy of the Expected Future

In this section, we propose our novel objective functional, which we call the free energy of the expected future (FEEF), which possesses the same epistemic value term as the EFE, while possessing a more naturalistic and intuitive grounding. We begin with the intuition that to act adaptively, agents should act so as to minimize the difference between what they predict will happen and what they desire to happen. Put another way, adaptive action for an agent consists of forcing reality to unfold according to its preferences. We can mathematically formulate this objective as the KL-divergence between the agent's veridical generative model of what is likely to happen and a biased generative model of what it desires to happen:

$$\pi^* = \underset{\pi}{\operatorname{argmin}} \mathbf{D}_{KL}[Q(o_{t:T}, x_{t:T}|\pi) || \tilde{p}(o_{t:T}, x_{t:T})].$$

The FEEF can be interpreted as the divergence between a veridical and a biased generative model, and thus furnishes a direct intuition of the goals of a FEEF-minimizing agent. The divergence objective compels the agent to bring the biased and the veridical generative model into alignment. Since the predictions of the biased generative model are heavily biased toward the agent's a priori preferences, the only way to achieve this alignment is to act so as to make the veridical generative model predict desired outcomes in line with the biased generative model. The FEEF objective encompasses the standard active inference intuition of an agent acting through biased inference to maximize accuracy of a biased model. However, the maintenance of two separate generative models (one biased and one veridical) also helps finesse the conceptual difficulty of how the agent manages to make accurate posterior inferences and future predictions about complex dynamics if all it has access to is a biased generative model. It seems straightforward that the biased model would also bias these crucial parts of inference that need to be unimpaired for the scheme to function at all. However, by keeping both a veridical generative model (the same one used at the present time and learned through environmental interactions) and a biased generative model (created by systematically biasing a temporary copy of the veridical

model), we elegantly separate the need for both veridical and biased inferential components for future prediction.⁹

Similar to the EFE, the FEEF objective can be decomposed into an extrinsic and an intrinsic term. We compare this directly to the EFE decomposition:

$$\begin{aligned} \text{FEEF}(\pi)_\tau &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} \left[\ln \left(\frac{Q(o_\tau, x_\tau | \pi)}{\tilde{p}(o_\tau, x_\tau)} \right) \right] \\ &= \underbrace{\mathbb{E}_{Q(x_\tau | \pi)} \mathbf{D}_{\text{KL}}[Q(o_\tau | x_\tau) \| \tilde{p}(o_\tau)]}_{\text{Extrinsic Value}} - \underbrace{\mathbb{E}_{Q(o_\tau | \pi)} \mathbf{D}_{\text{KL}}[Q(x_\tau | o_\tau) \| Q(x_\tau | \pi)]}_{\text{Intrinsic Value}}, \\ \text{EFE} &= \underbrace{-\mathbb{E}_{Q(o_\tau, x_\tau | \pi)} \left[\ln \tilde{p}(o_\tau) \right]}_{\text{Extrinsic Value}} - \underbrace{\mathbb{E}_{Q(o_\tau | \pi)} \mathbf{D}_{\text{KL}}[Q(x_\tau | o_\tau) \| Q(x_\tau | \pi)]}_{\text{Intrinsic Value}}. \end{aligned}$$

The first thing to note is that the intrinsic value terms of the FEEF and the EFE are identical, under the assumption that the variational posterior is approximately correct $Q(x_\tau | o_\tau) \approx p(x_\tau | o_\tau)$ such that FEEF-minimizing agents will necessarily show identical epistemic behavior to EFE-minimizing agents. Unlike the EFE, however, the FEEF also possesses a strong naturalistic grounding as a bound on a theoretically relevant quantity. The FEEF can maintain both its information-maximizing imperative and its theoretical grounding since it is derived from the minimization of a KL-divergence rather than the maximization of a log model evidence.

The key difference with the EFE lies in the likelihood term. While the EFE simply tries to maximize the expected evidence of the desired observations, the FEEF minimizes the KL-divergence between the likelihood of observations predicted under the veridical generative model¹⁰ and the marginal likelihood of observations under the biased generative model. This difference is effectively equivalent to an additional veridical generative model expected likelihood entropy term $\mathbf{H}[Q(o_\tau | x_\tau)]$ subtracted from the EFE. The extrinsic value term thus encourages the agent to choose its actions such that its predictions over states lead to observations that are close to its preferred observations, while also trying to move to states whereby the entropy over observations is maximized, thus leading the agent to move toward

⁹This approach bears a resemblance to that taken in Friston (2019), which separates the evolving dynamical policy-dependent density of the agent and a desired steady-state density that is policy invariant. This approach arises from deep thermodynamic considerations in continuous time, while ours is applicable to discrete time reinforcement learning frameworks.

¹⁰The term *veridical* needs some contextualizing. We simply mean that the model is not biased toward the agent's desires. The veridical generative model is not required to be a perfectly accurate map of the agent's entire world, only of action-relevant submanifolds of the total space (Tschantz, Seth et al., 2019).

states where the generative model is not as certain about the likely outcome. In effect, the FEEF possesses another exploratory term, in addition to the information gain, which the EFE lacks.

Another important advantage of the FEEF is that it is mathematically equivalent to the VFE (with a biased generative model) in the present time with a current observation. This is because when we have a real observation, the distribution over the possible veridical observations collapses to a delta distribution, so that the outer expectation has no effect as $\mathbb{E}_{Q(o_\tau, x_\tau | \pi)} = \int Q(x_\tau | o_\tau) Q(o_\tau | \pi) = \int Q(x_\tau | o_\tau) \delta(o - \bar{o}) = \int Q(x_\tau | \bar{o}_\tau)$ when a real observation \bar{o} is available. Similarly, the veridical model can be factorized as $Q(o_\tau, x_\tau) = Q(x_\tau | o_\tau) Q(o_\tau)$, and when the observation is known, the entropy of the observation marginal $Q(o_\tau | \pi)$ is 0, thus resulting in the VFE. Simultaneously, biased likelihood is equivalent to the veridical likelihood $\tilde{p}(\bar{o}_\tau | x_\tau) = Q(\bar{o}_\tau | x_\tau)$, assuming that (barring counterfactual reasoning capability) one cannot usefully desire things to be other than how they are at the present moment. This means that theoretically, we can consider an agent to be both inferring and planning using the same objective, which is not true of the EFE. The EFE does not reduce to the VFE when observations are known, and thus requires a separate objective function to be minimized for planning compared to perceptual inference. Because of this, it is possible to argue that FEEF is mandated by the free-energy principle. On this view, there is no distinction between present and future inference, and both follow from minimizing the same objective but under different informational constraints.

Since the FEEF and the EFE are identical in their intrinsic value term and share deep similarities in their extrinsic term, we believe that the FEEF can serve as a relatively straightforward “plug-in replacement” for the EFE for many active inference agents. Moreover, it has a much more straightforward intuitive basis than the EFE, is arguably a better continuation of the VFE into the future, and possesses a strong naturalistic grounding as a bound on the divergence between predicted and desired futures.

6 Discussion

We believe it is valuable at this point to step back from the morass of various free energies and take stock of what has been achieved. First, we have shown that it is not possible to directly derive epistemic value from variational inference objectives, which serve as a bound on model evidence. However, it is possible to derive epistemic value terms from divergences between the biased and veridical generative models. A deep intuitive understanding of why this is the case is an interesting avenue for future work. The intuition behind the FEEF as a divergence between desired and expected future observations is also similar to probabilistic formulations of the reinforcement learning problem (Attias, 2003; Kappen, 2005; Levine, 2018; Toussaint, 2009), which typically try to minimize the divergence between a controlled trajectory and an optimal trajectory (Kappen, 2007; Theodorou &

Todorov, 2012; Williams, Aldrich, & Theodorou, 2017). These schemes also obtain some degree of (undirected) exploratory behavior through their objective functionals, which contain entropy terms and the FEEF can be seen as a way of extending these schemes to partially observed environments. Understanding precisely how active inference and the free-energy principle relate mathematically to such schemes is another fruitful avenue for future work.

It seems intuitive that a Bayes-optimal solution to the exploration-exploitation dilemma should arise directly out of the formulation of reward maximization as inference, given that sources of uncertainty are correctly quantified. However, in this letter, we have shown that merely quantifying uncertainty in states and observations through mean-field-factorized time steps is insufficient to derive such a principled solution to the dilemma, as seen by the exploration-discouraging behavior of the FEF. We therefore believe that to derive Bayes-optimal exploration policies in the context of active learning, such that we have to select actions that give us the most information now to use in the future to maximize rewards, it is likely to require both modeling multiple interconnected time steps, as well as the mechanics of learning with parameters and update rules, and correctly quantifying the uncertainties therein. This is beyond the scope of this letter, but is a very interesting avenue for future work.

The comparison of the FEEF and the EFE also raises an interesting philosophical point about the number and types of generative models employed in the active-inference formalism. One interpretation of the FEEF is in terms of two generative models, but other interpretations are possible, such as between a single unbiased generative model and a simple density of desired states and observations. It is also important to note that due to requiring different objective functions for inference and planning, the EFE formulation also appears to implicitly require two generative models: the generative model of future states and the generative model of states in the future (Friston et al., 2015). While the mathematical formalism is relatively straightforward, the philosophical question of how to translate the mathematical objects into ontological objects called “generative models” is unclear, and progress on this front would be useful in determining the philosophical status, and perhaps even neural implementation of active inference.

The implications of our results for studies of active inference are varied. Nothing in what we have shown argues directly against the use of the EFE as an objective for an active inference agent. However, we believe we have shown that the EFE is not necessarily the only, or even the natural, objective function to use. We thus follow Biehl et al. (2018) in encouraging experimentation with different objective functions for active inference. We especially believe that our objective, the FEEF, has promise due its intuitive interpretation, largely equivalent terms to the EFE, its straightforward use of two generative models rather than just a single biased one, and its close connections to similar probabilistic objectives used in variational

reinforcement learning, while also maintaining the crucial epistemic properties of the EFE. Moreover, while in this letter we have argued for the FEF instead of the EFE as a direct extension of the VFE into the future, the logical requirements of exactly which functional (if any) is, in fact, mandated by the free-energy principle remains open. We believe that elucidating the exact constraints which the free-energy principle places on a theory of variational action, and understanding more deeply the relations between the various free energies, could shed light on deep questions regarding notions of Bayes-optimal epistemic action in self-organising systems.

Finally, it is important to note that although in this letter, we have solely been concerned with the EFE and active inference in discrete-time POMDPs, the original intuitions and mathematical framework of the free-energy principle arose out of a continuous time formulation, deeply interwoven with concerns from information theory and statistical physics (Friston, 2019; Friston & Ao, 2012; Friston et al., 2006; Parr et al., 2020). As such, there may be deep connections between the EFE, FEF, and log model evidence that exist only in the continuous time limit and that furnish a mathematically principled origin of epistemic action.

7 Conclusion

In this letter, we have examined in detail the nature and origin of the EFE. We have shown that it is not a direct analog of the VFE extended into the future. We then derived a novel objective, the FEF, which we claimed is a more natural extension and shown that it lacks the beneficial epistemic value term of the EFE. We then proved that this term arises in the EFE directly as a result of its nonstandard definition since the EFE can be expressed as just the FEF minus the expected information gain. Taking this into account, we then proposed another objective, the free energy of the expected future (FEFF), which attempts to get the best of both worlds by preserving the desirable information-seeking properties of the EFE, while also maintaining a mathematically principled origin.

Appendix A: Variational Inference

To motivate the variational free energy, and variational inference more generally, we set up a standard inference problem. Let us say we are an agent that exists in a partially observed world. We have some observation o_t , and from this, we wish to infer the hidden state of the world x_t . That is, we want to compute the posterior $p(x_t|o_t)$. While we do not know this posterior directly, we do possess a generative model of the world. This is a model that maps from hidden states to observations. Mathematically, we possess $p(o_t, x_t) = p(o_t|x_t)p(x_t)$. Since computing the true posterior exactly is likely intractable, the strategy in variational inference is to try to approximate this density with a tractable one $Q(x_t|o_t; \phi)$, which we postulate,

and thus have full control over. While the true posterior might be arbitrarily complex, we might define $Q(x_t|o_t; \phi)$ to be a gaussian distribution: $Q(x_t|o_t; \phi) = \mathcal{N}(x; \mu_\phi, \sigma_\phi)$, for instance. Given that we have this variational density q , parameterized by some parameters ϕ , the goal is to adjust the parameters to make q as close as possible to the true posterior $p(x_t|o_t)$. Mathematically, this means we want to minimize

$$\operatorname{argmin}_\phi \mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(x_t|o_t)],$$

where $\mathbf{D}_{KL}[Q||P]$ is the KL-divergence. This initially doesn't seem to have bought us much. We wish to minimize the divergence between the variational density q and the true posterior $p(x_t|o_t)$. However, by assumption, we do not know the true posterior. So how can we possibly minimize this divergence if we do not know one of the parts? This is where we use the key trick of variational inference. By Bayes' theorem, we know that $p(x_t|o_t) = \frac{p(o_t|x_t)p(x_t)}{p(o_t)}$ and we can thus substitute this into the KL divergence term:

$$\begin{aligned} & \operatorname{argmin}_\phi \mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(x_t|o_t)] \\ &= \operatorname{argmin}_\phi \mathbf{D}_{KL}\left[Q(x_t|o_t; \phi)||\frac{p(o_t|x_t)p(x_t)}{p(o_t)}\right], \\ &= \mathbb{E}_{Q(x_t|o_t; \phi)} \left[\ln \left(\frac{Q(x_t|o_t; \phi)p(o_t)}{p(o_t|x_t)p(x_t)} \right) \right], \\ &= \mathbb{E}_{Q(x_t|o_t; \phi)} \left[\ln \left(\frac{Q(x_t|o_t; \phi)}{p(o_t|x_t)p(x_t)} \right) \right] + \mathbb{E}_{Q(x_t|o_t; \phi)} [\ln p(o_t)] \\ &= \mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(o_t|x_t)p(x_t)] + \ln p(o_t). \end{aligned} \tag{A.1}$$

In step 2 we applied Bayes' theorem to the posterior. In step 3 we simply utilized the definition of the KL divergence $\mathbf{D}_{KL}[Q||P] = \mathbb{E}_Q \ln \left(\frac{Q}{P} \right)$. In step 4 we applied the property of logs that $\ln(a * b) = \ln(a) + \ln(b)$. In step 5 we recognize that the remaining first term is now a KL-divergence between the variational posterior and the generative model. We also recognize that since the $\ln p(o_t)$ term has no dependence on x or ϕ , the expectation $\mathbb{E}_{Q(x_t|o_t; \phi)} \ln p(o_t)$ vanishes, leaving just the $\ln p(o_t)$ term alone. It is important to note that the KL term in equation A.1 is now between two things we can actually compute: the variational posterior, which we control, and the generative model, which we assume that we know. The remaining $\ln p(o_t)$ term is called the log model evidence and is incomputable in general. However, since it is not affected by the parameters ϕ of the variational density, it does not affect the minimization, and so for the purposes of the minimization

process can be ignored. We can thus write out what we have defined as

$$\begin{aligned} \mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(x_t|o_t)] &= \mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(o_t|x_t)p(x_t)] + \ln p(o_t) \\ \Rightarrow \mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(o_t|x_t)p(x_t)] &\geq \mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(x_t|o_t)]. \end{aligned}$$

This implies that the KL-divergence between the variational density and the generative model is always greater than or equal to the KL-divergence between the true and variational posteriors. Since we can compute the first KL-divergence, we call it the variational free energy F . Since it is an upper bound on the divergence between the true posterior and the variational posterior, which is what we really want to minimize, then if we minimize F , we are constantly pushing that bound lower and thus largely minimizing the divergence between the true and variational posterior. As an additional bonus, when the true and variational posteriors are approximately equal, $\mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(x_t|o_t)] \approx 0$ then $\mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(o_t|x_t)p(x_t)] \approx -\ln p(o_t)$, which means that the final value of the variational free energy is thus equal to the negative log model evidence. Since the log model evidence is a very useful quantity to compute for Bayesian model selection, it effectively means that once we have finished fitting our model, we are automatically left with a measure of how good our model is.

In effect, the variational free energy is useful because it has two properties. The first is that it is an upper bound on the divergence between the true and approximate posterior. By adjusting our approximate posterior to minimize this bound, we drive it closer to the true posterior, thus achieving more accurate inference. Second, the variational free energy is a bound on the log model evidence. This is an important term that scores the likelihood of the data observed given the model and so can be used in Bayesian model selection.

The log model evidence takes on additional importance in terms of the free-energy principle, since the negative log model evidence $-\ln p(o_t)$ is surprisal, which all agents, it is proposed, are driven to minimize (Friston et al., 2006). This is because the expected log model evidence is the entropy of observations, the minimization of which is postulated as a necessary condition for any self-sustaining organism to maintain itself as a unique system. The free-energy minimization comes about since the VFE is, as we have seen, a tractable bound on the log model evidence, or surprisal.

The VFE can be decomposed in three principal ways, each showcasing a different facet of the objective:

$$\begin{aligned} \mathbb{F} &= \mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(o_t, x_t)] \\ &= \mathbb{E}_{Q(x_t|o_t; \phi)} \left[\ln \frac{Q(x_t|o_t; \phi)}{p(o_t, x_t)} \right] \end{aligned}$$

$$\begin{aligned}
 &= \underbrace{\mathbb{E}_{Q(x_t|o_t; \phi)}[\ln Q(x_t|o_t; \phi)]}_{\text{Entropy}} - \underbrace{\mathbb{E}_{Q(x_t|o_t; \phi)}[\ln p(o_t, x_t)]}_{\text{Energy}} \\
 &= \underbrace{-\mathbb{E}_{Q(x_t|o_t; \phi)}[\ln p(o_t|x_t)]}_{\text{Accuracy}} + \underbrace{\mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(x_t)]}_{\text{Complexity}} \\
 &= \underbrace{-\ln p(o_t)}_{\text{Negative Log Model Evidence}} + \underbrace{\mathbf{D}_{KL}[Q(x_t|o_t; \phi)||p(x_t|o_t)]}_{\text{Posterior Divergence}}.
 \end{aligned}$$

In the first entropy-energy decomposition, we simply split the KL-divergence using the properties of logarithms so that the numerator of the fraction becomes the entropy term and the denominator becomes the energy term. If we are seeking to minimize the variational free energy, we need to minimize both the negative entropy (since entropy is defined as $-\mathbb{E}_{Q(x)}[\ln Q(x)]$) and the negative energy (or maximize the energy) $\mathbb{E}_{Q(x_t|o_t; \phi)}[\ln p(o_t, x_t)]$. This can be interpreted as saying we require that the variational posterior be as entropic as possible while also maximizing the likelihood that the x s proposed as probable by the variational posterior also be judged as probable under the generative model.

The second decomposition into accuracy and complexity perhaps has a more straightforward interpretation. We wish to minimize the negative accuracy (and thus maximize the accuracy), which means we want the observed observation to be as likely as possible under the x s predicted by the variational posterior. However, we also want to minimize the complexity term, which is a KL-divergence between the variational posterior and the prior. That is, we wish to keep your posterior as close to our prior as possible while still maximizing accuracy. The complexity term then functions as a kind of implicit regularizer, making sure we do not overfit to any specific observation.

The final decomposition speaks to the inferential functions of the VFE. It serves as an upper bound on the log model evidence, since the posterior divergence term, as a KL-divergence, is always positive. Moreover, we see that by minimizing the free energy, we must also be minimizing the posterior divergence, which is the difference between the approximate and true posterior, and we are thus improving our variational approximation.

Appendix B: Decompositions of the EFE

In this section we provide a comprehensive overview of the many decompositions of the EFE. The EFE is defined as

$$\mathcal{G}(\pi) = \mathbb{E}_{Q(o_\tau, x_\tau|\pi)}[\ln Q(x_\tau|\pi) - \ln \tilde{p}(o_\tau, x_\tau)].$$

The standard decomposition is into the extrinsic term (expected log likelihood of the desired observations) and an epistemic term (the information gain, or KL-divergence between variational prior and posterior from the generative model):

$$\begin{aligned} & \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau, x_\tau)] \\ &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [-\ln \tilde{p}(o_\tau) - \ln p(x_\tau | o_\tau) + \ln Q(x_\tau | \pi)] \\ &= \underbrace{-\mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln \tilde{p}(o_\tau)]}_{\text{Extrinsic Value}} - \underbrace{\mathbb{E}_{Q(o_\tau | \pi)} [\mathbf{D}_{KL}[Q(x_\tau | o_\tau) \| Q(x_\tau | \pi)]]}_{\text{Epistemic Value}}. \end{aligned}$$

Similar to the VFE, it is also possible to split it into an energy and an entropy term. While the energy term is similar to the VFE as the expectation of the generative model (albeit an expectation under the joint instead of the posterior), the entropy term is different as it is the entropy of the variational prior, not the approximate posterior, which results:

$$\begin{aligned} \mathcal{G}(\pi) &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau, x_\tau)] \\ &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi)] - \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln \tilde{p}(o_\tau, x_\tau)] \\ &= \underbrace{-\mathbb{E}_{Q(o_\tau | x_\tau)} [\mathcal{H}[Q(x_\tau | \pi)]]}_{\text{Entropy}} - \underbrace{\mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln \tilde{p}(o_\tau, x_\tau)]}_{\text{Energy}}. \end{aligned}$$

It is also possible to decompose the biased generative model the other way around, thus in line with that of the VFE to derive

$$\begin{aligned} \mathcal{G}(\pi) &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau, x_\tau)] \\ &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau | x_\tau) - \ln p(x_\tau)] \\ &= \underbrace{-\mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln \tilde{p}(o_\tau | x_\tau)]}_{\text{Accuracy}} + \underbrace{\mathbb{E}_{Q(o_\tau | x_\tau)} [\mathbf{D}_{KL}[Q(x_\tau | \pi) \| p(x_\tau)]]}_{\text{Complexity}}. \end{aligned}$$

Unlike the VFE, however, the divergence is between the variational prior and the generative prior rather than between the variational posterior and the generative prior. Finally, the EFE can also be represented in observation space by using Bayes' rule to flip the likelihoods and priors:

$$\begin{aligned} \mathcal{G}(\pi) &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau, x_\tau)] \\ &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau) - \ln Q(x_\tau | o_\tau)] \\ &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau) - \ln Q(o_\tau | x_\tau) \\ &\quad - \ln Q(x_\tau | \pi) + \ln Q(o_\tau)] \end{aligned}$$

$$\begin{aligned}
 &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [-\ln \tilde{p}(o_\tau) - \ln Q(o_\tau | x_\tau) + \ln Q(o_\tau)] \\
 &= \underbrace{\mathbb{E}_{Q(x_\tau | \pi)} [\mathcal{H}[Q(o_\tau | x_\tau)]]}_{\text{Predicted Uncertainty}} - \underbrace{\mathbb{E}_{Q(x_\tau | o_\tau)} [\mathbf{D}_{KL}[Q(o_\tau) \| \tilde{p}(o_\tau)]]}_{\text{Predicted Divergence}} \\
 &= - \underbrace{\mathbb{E}_{Q(x_\tau | \pi)} [\ln \tilde{p}(o_\tau)]}_{\text{Extrinsic Value}} - \underbrace{\mathbb{E}_{Q(x_\tau | \pi)} [\mathbf{D}_{KL}[Q(o_\tau | x_\tau) \| Q(o_\tau)]]}_{\text{(Observation) Information Gain}}.
 \end{aligned}$$

It is also possible to factorize the biased generative model the other way around in terms of an unbiased likelihood and biased states: $\tilde{p}(o_\tau, x_\tau) = p(o_\tau | x_\tau) \tilde{p}(x_\tau)$. This different factorization leads to a new decomposition in terms of risk and ambiguity, as well as potentially different behavior due to the change from desired observations to desired states:¹¹

$$\begin{aligned}
 \mathcal{G}(\pi) &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau, x_\tau)] \\
 &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln p(o_\tau | x_\tau) - \ln \tilde{p}(x_\tau)] \\
 &= \underbrace{\mathbb{E}_{Q(x_\tau | \pi)} [\mathcal{H}[p(o_\tau | x_\tau)]]}_{\text{Ambiguity}} + \underbrace{\mathbf{D}_{KL}[Q(x_\tau | \pi) \| \tilde{p}(x_\tau | \pi)]}_{\text{Risk}}.
 \end{aligned}$$

Here the agent is driven to minimize the divergence between desired and prior expected states, while also trying to minimize the entropy of the observations it receives. This drives the agent to try to sample observations with a minimally ambiguous (or maximally precise) mapping back to states.

This formulation is mathematically equivalent to the previous decompositions despite defining desired states instead of desired observations, as can be seen with the following manipulations:

$$\begin{aligned}
 \mathcal{G}(\pi) &= \underbrace{\mathbb{E}_{Q(x_\tau | \pi)} [\mathcal{H}[p(o_\tau | x_\tau)]]}_{\text{Ambiguity}} + \underbrace{\mathbf{D}_{KL}[Q(x_\tau) \| \tilde{p}(x_\tau)]}_{\text{Risk}} \\
 &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln p(o_\tau | x_\tau) - \ln Q(x_\tau | o_\tau) - \ln \tilde{p}(o_\tau) \\
 &\quad + \ln p(o_\tau | x_\tau)] \\
 &= - \underbrace{\mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln \tilde{p}(o_\tau)]}_{\text{Extrinsic Value}} - \underbrace{\mathbb{E}_{Q(o_\tau | \pi)} [\mathbf{D}_{KL}[Q(x_\tau | o_\tau) \| Q(x_\tau | \pi)]]}_{\text{Epistemic Value}} = \mathcal{G}(\pi).
 \end{aligned}$$

¹¹ For further detail on this factorization see Da Costa et al. (2020).

The risk-ambiguity formulation has very close relations to KL control (Rawlik et al., 2013), in that it encompasses KL control with an additional “epistemic” ambiguity term;

$$\mathcal{G}(\pi) = \underbrace{\mathbb{E}_{Q(x_\tau|\pi)} \mathcal{H}[p(o_\tau|x_\tau)]}_{\text{Active Inference}} + \underbrace{\mathbf{D}_{KL}[Q(x_\tau|\pi) \parallel \tilde{p}(x_\tau)]}_{\text{KL Control}}.$$

Appendix C: Trajectory Derivation of the Expected Model Evidence

Here we present the derivation of the free energy of the future (FEF) from the expected model evidence for the full trajectory distribution rather than a single time step. Importantly, we show that with a temporal mean-field approximation on the approximate posterior: $p(x_{1:T}|o_{1:T}) \approx \prod_t p(x_t|o_t)$, the assumption that desired rewards are independent in time: $\tilde{p}(o) \approx \prod_t^T p(\hat{r}_t)$, and given a Markovian generative model, then the trajectory distribution factorizes into a sum of individual time-steps,¹² only dependent on the past through the prior term $p(x_t) = \mathbb{E}_{Q(x_{t-1}|o_{t-1})} p(x_t|x_{t-1})$. We name this final approximation the factorization approximation, and it simply states that the prior at the current time step is based on the posterior of the previous time step mapped through the transition dynamics $p(x_t|x_{t-1})$:

$$\begin{aligned} & \operatorname{argmin}_{p(\pi)} -\mathbb{E}_{Q(o_{1:T}|\pi)} \ln \tilde{p}(o_{1:T}) \\ &= -\mathbb{E}_{Q(o_{1:T}|\pi)} \left[\ln \int dx_{1:T} \tilde{p}(o_{1:T}, x_{1:T}) \right] \\ &= -\mathbb{E}_{Q(o_{1:T}|\pi)} \left[\ln \int dx_{1:T} \frac{\tilde{p}(o_{1:T}, x_{1:T}) Q(x_{1:T}|o_{1:T})}{Q(x_{1:T}|o_{1:T})} \right] \\ &= -\mathbb{E}_{Q(o_{1:T}|\pi)} \left[\ln \int dx_{1:T} \prod_t^T \frac{\tilde{p}(o_t, x_t) Q(x_t|o_t)}{Q(x_t|o_t)} \right] \\ &= -\mathbb{E}_{Q(o_{1:T}|\pi)} \left[\ln \int dx_{1:T} \prod_t^T \frac{\tilde{p}(o_t|x_t) \mathbb{E}_{Q(x_{t-1}|o_{t-1})} p(x_t|x_{t-1}) Q(x_t|o_t)}{Q(x_t|o_t)} \right] \\ &= \mathbb{E}_{Q(o_{1:T}|\pi)} \left[\sum_t^t \ln \int dx_t \frac{\tilde{p}(o_t|x_t) - \mathbb{E}_{Q(x_{t-1}|o_{t-1})} p(x_t|x_{t-1}) Q(x_t|o_t)}{Q(x_t|o_t)} \right] \\ &\geq -\sum_t^t \mathbb{E}_{Q(o_{1:T}|\pi)} \left[\int dx_t Q(x_t|o_t) \ln \frac{\tilde{p}(o_t|x_t) \mathbb{E}_{Q(x_{t-1}|o_{t-1})} p(x_t|x_{t-1})}{Q(x_t|o_t)} \right] \end{aligned}$$

¹²We assume discrete time so there is a sum over time steps. We also assume continuous states so there is an integral over states x . However, the derivation is identical in the case of discrete states where the integral is simply replaced with a sum.

$$\begin{aligned}
 &\geq - \sum_t^t \int dx_t \int do_{1:T} Q(o_{1:T}, x_t | \pi) \ln \frac{\tilde{p}(o_t | x_t) \mathbb{E}_{Q(x_{t-1} | o_{t-1})} p(x_t | x_{t-1})}{Q(x_t | o_t)} \\
 &\geq - \sum_t^t \mathbb{E}_{Q(o_t, x_t | \pi)} \left[\ln \frac{\tilde{p}(o_t | x_t) \mathbb{E}_{Q(x_{t-1} | o_{t-1})} p(x_t | x_{t-1})}{Q(x_t | o_t)} \right] \\
 &\geq - \sum_t^t \mathbb{E}_{Q(o_t, x_t | \pi)} \left[\ln \tilde{p}(o_t | x_t) - \sum_t^t \mathbb{E}_{p(o_t)} \mathbf{D}_{KL}[p(x_t | o_t) || \mathbb{E}_{Q(x_{t-1} | o_{t-1})} p(x_t | x_{t-1})] \right] \\
 &\geq - \sum_t^t \mathbf{FEF}_t.
 \end{aligned}$$

The trajectory derivation of the FEEF follows an almost identical scheme to that of the FEF. The only difference is that now the term inside the log also contains an additional $-\ln \tilde{p}(o)$, which is then combined with the likelihood from the generative model to form the extrinsic-value KL-divergence.

Appendix D: EFE Bound on the Negative Log Model Evidence _____

It is important to note that the EFE is also a bound on the negative log model evidence, but a lower bound, not an upper bound. This means that in theory, one should want to maximize the EFE, instead of minimize it, to make the bound as tight as possible.

It is straightforward to show the bound, since the extrinsic value term of the EFE simply *is* the log model evidence:

$$\begin{aligned}
 \mathbf{EFE} &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau, x_\tau)] \\
 &\approx \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln Q(x_\tau | o_\tau) - \ln \tilde{p}(o_\tau)] \\
 &\approx \underbrace{-\mathbb{E}_{Q(o_\tau | \pi)} [\ln \tilde{p}(o_\tau)]}_{\text{Negative Expected Log Model Evidence}} \quad - \underbrace{\mathbb{E}_{Q(o_\tau | \pi)} \mathbf{D}_{KL}[Q(x_\tau | o_\tau) || Q(x_\tau | \pi)]}_{\text{Information Gain}}.
 \end{aligned}$$

This derivation assumes that the true and approximate posteriors are approximately equal, $p(x_\tau | o_\tau) \approx Q(x_\tau | o_\tau)$, such that this is true only after a variational inference procedure is completed.

We wish to minimize both log model evidence and minimize the EFE. Since the information gain term is a KL-divergence, which is always ≥ 0 , and we have a negative information gain term, this means that the EFE is always less than the log model evidence and so is a lower bound. However, this bound becomes tight when the information gain is 0, so to maximally tighten the bound, we wish to reduce the information gain while the EFE demands we maximize it. In effect, this means that the EFE bound is the wrong way around.

We can see this more clearly when we retrace the logic for the FEF. From equation 4.1, we have that the FEF is an upper bound on the negative log model evidence. This means that minimizing the FEF necessarily tightens the bound, while this is not true of the EFE lower bound, where minimizing the EFE can actually cause it to diverge from the log model evidence. We can see this even more clearly by doing an analogous decomposition of the FEF:

$$\begin{aligned} \text{FEF} &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | o_\tau) - \ln \tilde{p}(o_\tau, x_\tau)] \\ &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | o_\tau) - \ln p(x_\tau | o_\tau) - \ln \tilde{p}(o_\tau)] \\ &= \underbrace{-\mathbb{E}_{Q(o_\tau | \pi)} [\ln \tilde{p}(o_\tau)]}_{\text{Negative Expected Log Model Evidence}} + \underbrace{\mathbb{E}_{Q(o_\tau | \pi)} \mathbf{D}_{KL}[Q(x_\tau | o_\tau) \| p(x_\tau | o_\tau)]}_{\text{Posterior Approximation Error}}. \end{aligned}$$

Here, since the KL is between the generative model and the approximate posterior, and then decompose the generative model into a true posterior and marginal, we can no longer make the assumption, made in the EFE derivation, that the true and approximate posterior are approximately equal, since that would leave us with only the model evidence. Therefore, instead we get a posterior approximation error term, which is the KL-divergence between the approximate and true posteriors. When the true and approximate posteriors are equal, we are left with the log model evidence. Since the posterior approximation error is always ≥ 0 , the FEF is an upper bound on the negative log model evidence, and thus by minimizing the FEF, we make the bound tighter. This logic is essentially a reprise of the standard variational inference logic from a slightly different perspective.

If we do not make the assumption in the EFE that the approximate and true posterior are the same, we can derive a similar expression to the EFE that will shed more light on the relation:

$$\begin{aligned} \text{EFE} &= \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln \tilde{p}(o_\tau, x_\tau)] \\ &\approx \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln p(x_\tau | o_\tau) - \ln \tilde{p}(o_\tau)] \\ &\approx \mathbb{E}_{Q(o_\tau, x_\tau | \pi)} [\ln Q(x_\tau | \pi) - \ln p(x_\tau | o_\tau) - \ln \tilde{p}(o_\tau) + \ln Q(x_\tau | o_\tau) \\ &\quad - \ln Q(x_\tau | o_\tau)] \\ &\approx \underbrace{-\mathbb{E}_{Q(o_\tau | \pi)} [\ln \tilde{p}(o_\tau)]}_{\text{Negative Expected Log Model Evidence}} + \underbrace{\mathbb{E}_{Q(o_\tau | \pi)} \mathbf{D}_{KL}[Q(x_\tau | o_\tau) \| p(x_\tau | o_\tau)]}_{\text{Posterior Approximation Error}} \\ &\quad \underbrace{\hspace{10em}}_{\text{FEF}} \\ &\quad - \underbrace{\mathbb{E}_{Q(o_\tau | \pi)} \mathbf{D}_{KL}[Q(x_\tau | o_\tau) \| Q(x_\tau | \pi)]}_{\text{Information Gain}}. \end{aligned}$$

Without the true posterior assumption, we thus find that the EFE could be an upper or a lower bound on the log model evidence, since the two additional KL-divergence terms have opposite signs. If the posterior approximation error is larger than the information gain, the EFE functions correctly as an upper bound. However, if the information gain is larger, the EFE will become a lower bound and could diverge from the log model evidence. This latter situation is more likely since the goal of variational inference is to reduce the approximation error, while EFE agents seek to maximize information gain. This means that the EFE functions correctly as an upper bound on log model evidence only during the early stages of optimization when the posterior approximation is poor. Further optimization steps likely drive the EFE further away from the model evidence. The bound is tight when the information gain equals the posterior approximation error. We can also see that the first two terms of the EFE are simply the FEF; we have thus red-erived by a rather round-about route the fact that the EFE is simply the FEF minus the information gain.

We thus see that the status of the EFE as a bound on the log model evidence is shaky, since it depends on the information gain always being larger or smaller than the posterior approximation error. Moreover, the bounding behavior seems to emerge directly from the relation of the EFE to the FEF rather than the intrinsic qualities of the EFE, and it is primarily the information-seeking properties of the EFE that serve to damage the clean bounding behavior of the FEF.

It can be argued that although the mathematical justification of the EFE as a bound may be shaky, the additional information gain term may be beneficial, and the bound may be recovered in the long run, since as a result of short-term actions to maximize the EFE, the epistemic value itself goes to 0, and thus the EFE exactly approximates the bound, while also potentially increasing the ultimate expected reward achieved. This argument is valid heuristically and is identical to the standard justifications for ad hoc intrinsic measures terms in the literature (Oudeyer & Kaplan, 2009)—namely, that exploration hurts in the short run but helps in the long run. We do not dispute that argument in this letter; instead we simply show that the EFE cannot straightforwardly be justified mathematically as being a result of variational inference into the future or as a bound on model evidence. We do not argue at all against its heuristic use to encourage exploration of the environment and thus (we hope) better performance overall.

Appendix E: Attempts at Naturalizing the EFE _____

In this appendix, we review several attempts to derive the EFE directly from the expected model evidence.

Since we have derived the FEF by importance-sampling the expected model evidence with the approximate posterior, one obvious avenue would be to importance-sample on the variational prior instead. Following this

line of thought gives us

$$\begin{aligned}
 -\mathbb{E}_{Q(o_\tau|\pi)}[\ln \tilde{p}(o_\tau)] &= -\mathbb{E}_{Q(o_\tau|\pi)}\left[\ln \int dx_\tau \tilde{p}(o_\tau, x_\tau)\right] \\
 &= -\mathbb{E}_{Q(o_\tau|\pi)}\left[\ln \int dx_\tau \tilde{p}(o_\tau, x_\tau) \frac{Q(x_\tau|\pi)}{Q(x_\tau|\pi)}\right] \\
 &\leq -\mathbb{E}_{Q(o_\tau|\pi)}\left[\int dx_\tau Q(x_\tau|\pi) \ln \frac{\tilde{p}(o_\tau, x_\tau)}{Q(x_\tau|\pi)}\right] \\
 &\leq -\mathbb{E}_{Q(o_\tau|\pi)Q(x_\tau|\pi)}\left[\ln \frac{\tilde{p}(o_\tau, x_\tau)}{Q(x_\tau|\pi)}\right] \\
 &\leq -\mathbb{E}_{Q(o_\tau|\pi)Q(x_\tau|\pi)}[\ln Q(x_\tau|\pi) - \ln \tilde{p}(o_\tau, x_\tau)].
 \end{aligned}$$

While this approach gets the correct form of the EFE inside the expectation, the expectation itself is the product of the two marginals rather than the joint required for the full EFE. While this may seem minor, this difference must underpin all the other differences and relations we have explored throughout this letter.

To get to the full EFE, we must make some assumptions to allow us to combine the expectation under two marginals into an expectation under the joint. The first and simplest assumption is that they are the same, such that the joint factorizes into the two marginals: $Q(o_\tau, x_\tau|\pi) \approx Q(o_\tau|\pi)Q(x_\tau|\pi)$. This assumption is equivalent to assuming independence of observations and latent states, which rather defeats the point of a latent variable model.

A second approach is to assume that the variational prior equals the variational posterior $Q(x_\tau|\pi) \approx Q(x_\tau|o_\tau)$. This allows one to combine the marginal and posterior into a joint, giving the EFE as desired. However this assumption has several unfortunate consequences. First, it eliminates the entire idea of inference, since the prior and posterior are assumed to be the same; thus, no real inference can have taken place. This is not necessarily an issue if we separate the inference and planning stages of the algorithm, such that they optimize different objective functions; however, the FEEF approach is more elegant as it enables the optimization of the same objective function for both inference and planning, thus casting them as different facets of the same underlying process. Moreover, a more serious issue is that this assumption also eliminates the information gain term in active inference; since the prior and posterior are the same, the divergence between them (which is the information gain), must be zero.

A slightly different approach is taken in a proof in Parr (2019), which begins with the KL-divergence between two distributions, one encoding beliefs about future states and observations and the other being the biased generative model. By definition, this KL-divergence is always ≥ 0 , which

allows us to write

$$\begin{aligned}
 & \mathbf{D}_{KL}[p(o_\tau, x_\tau | \pi) \| \tilde{p}(o_\tau, x_\tau)] \geq 0 \\
 & = \mathbb{E}_{p(o_\tau | \pi)} \mathbf{D}_{KL}[p(x_\tau | o_\tau) \| \tilde{p}(o_\tau, x_\tau)] - \mathbb{E}_{p(o_\tau | \pi)} [\ln p(o_\tau | \pi)] \geq 0 \\
 & \Rightarrow -\mathbb{E}_{p(o_\tau | \pi)} \mathbf{D}_{KL}[p(x_\tau | o_\tau) \| \tilde{p}(o_\tau, x_\tau)] \geq -\mathbb{E}_{p(o_\tau | \pi)} [\ln p(o_\tau | \pi)] \\
 & \Rightarrow \mathbf{FEF} \geq -\mathbb{E}_{p(o_\tau | \pi)} [\ln p(o_\tau | \pi)].
 \end{aligned}$$

Under the assumption that $p(x|o) \approx Q(x|\pi)$, this becomes

$$\begin{aligned}
 & -\mathbb{E}_{p(o_\tau | \pi)} \mathbf{D}_{KL}[p(x_\tau | o_\tau) \| \tilde{p}(o_\tau, x_\tau)] \geq -\mathbb{E}_{p(o_\tau | \pi)} [\ln p(o_\tau | \pi)] \\
 & \approx \mathbb{E}_{p(o_\tau | \pi)} \mathbf{D}_{KL}[Q(x_\tau | \pi) \| \tilde{p}(o_\tau, x_\tau)] \geq -\mathbb{E}_{p(o_\tau | \pi)} [\ln p(o_\tau | \pi)] \\
 & \approx \mathbf{EFE} \geq -\mathbb{E}_{p(o_\tau | \pi)} [\ln p(o_\tau | \pi)].
 \end{aligned}$$

This proof derives the FEF as a bound on not the expected model evidence by our definition, but on the entropy of expected observations given a policy. The EFE is then derived from the FEF by assuming that the prior and posterior are the same, which comes with all the drawbacks explained above. This proof is primarily unworkable because of the assumption that the prior and the posterior are identical. While this may be arguable in the continuous time limit, where it is equivalent to the assumption that that $\frac{dQ(x|o)}{dt} \approx 0$, which is when the continuous-time inference has reached an equilibrium, it is definitely not true in discrete time; although there is a relation between the prior in the current time step and the posterior in the previous one, it must be mapped through the transition dynamics – $Q(x_t | \pi) = \mathbb{E}_{Q(x_{t-1} | \pi)} [p(x_t | x_{t-1}, \pi)]$.

One can also attempt a related proof by splitting the KL-divergence the other way. This gives

$$\begin{aligned}
 & \mathbf{D}_{KL}[p(o_\tau, x_\tau | \pi) \| \tilde{p}(o_\tau, x_\tau)] \geq 0 \\
 & = \mathbf{D}_{KL}[p(o_\tau, x_\tau | \pi) \| \tilde{p}(x_\tau | o_\tau)] - \mathbf{E}_{p(o_\tau | \pi)} [\ln \tilde{p}(o_\tau)] \geq 0 \\
 & \Rightarrow -\mathbf{D}_{KL}[p(o_\tau, x_\tau | \pi) \| \tilde{p}(x_\tau | o_\tau)] \geq -\mathbf{E}_{p(o_\tau | \pi)} [\ln \tilde{p}(o_\tau)] \\
 & \Rightarrow \mathbf{E}_{p(x_\tau | \pi)} [\ln p(o_\tau | x_\tau)] + \mathbf{E}_{p(o_\tau | \pi)} \mathbf{D}_{KL}[p(x_\tau | o_\tau) \| \tilde{p}(x_\tau | \pi)] \geq -\mathbf{E}_{p(o_\tau | \pi)} [\ln \tilde{p}(o_\tau)] \\
 & \Rightarrow \mathbf{FEF} \geq -\mathbf{E}_{p(o_\tau | \pi)} [\ln \tilde{p}(o_\tau)],
 \end{aligned}$$

which is just another way of showing that the FEF is a bound on the expected model evidence.

Appendix F: Related Quantities

Recently a new free energy, the generalized free energy (GFE) (Parr & Friston, 2019), has been proposed in the literature as an alternative or an

extension to the EFE. The GFE shares some close similarities with the FEEF. Both fundamentally extend the EFE by proposing a unified objective function, which is valid for both inference at the current time and planning into the future, whereas the EFE can only be used for planning. Moreover, both GFE and FEEF encode future observations as latent unobserved variables, over which posterior beliefs can be formed. Moreover agents maintain prior beliefs over these variables, which encode its preferences or desires.¹³

The generalized free energy is defined as

$$\text{GFE} = \mathbb{E}_{Q(o_\tau, x_\tau)}[\ln Q(o_\tau) + \ln Q(x_\tau) - \ln \tilde{p}(o_\tau, x_\tau)],$$

whereas the FEEF is defined as

$$\text{FEEF} = \mathbb{E}_{Q(o_\tau, x_\tau)}[\ln Q(o_\tau, x_\tau) - \ln \tilde{p}(o_\tau, x_\tau)].$$

There are two key differences mathematically and intuitively between the GFE and the FEEF. The first is that the GFE maintains a factorized posterior over beliefs and observations, where the posterior beliefs of the two are separated by a mean-field approximation and assumed to be separate. By contrast, the FEEF maintains a joint approximate belief over both observations and states simultaneously. This joint in the case of the FEEF effectively functions as a veridical generative model since $Q(o|x) = p(o|x)$ and $Q(x) = \mathbb{E}_{Q(x_{t-1}|\pi)}p(x_t|x_{t-1})$. This means that posterior beliefs of the future are computed simply by rolling forward the generative model given the beliefs about the current time.

A second and more important differences lies in the generative models. The GFE assumes that the agent is only equipped with a single generative model with both veridical and biased components. The preferences of an EFE agent are encoded as a separate factorizable marginal over observations. This means that the generative model of the GFE agent factorizes as $\tilde{p}(o, x)_{GFE} \propto p(o|x)p(x)\tilde{p}(o)$. This means that for the GFE, the likelihood and the prior are unbiased, and there is simply an additional prior preferences term in the free-energy expression. By contrast, the FEEF eschews this unusual factorization of the generative model and instead presupposes a separate warped generative model for use in the future that is intrinsically biased. The FEEF generative model thus decomposes as $\tilde{p}(o, x)_{FEEF} = \tilde{p}(o|x)\tilde{p}(x)$, which is the standard factorization of the joint distribution in a generative model, but where both the likelihood and prior

¹³To help make clear the similarity between the GFE and the FEEF, we have defined the veridical generative model as $Q(o_\tau, x_\tau)$.

distributions are biased toward generating more favorable states of affairs for the agent. This inherent optimism bias then drives action.

A further free energy proposed in the literature has been the Bethe free energy and the Bethe approximation (Schwöbel et al., 2018). This approach eschews the standard mean-field assumption on the approximate posterior in favor of a Bethe approximation from statistical physics (Yedidia, Freeman, & Weiss, 2001, 2005), which instead represents the approximate posterior as the product of pairwise marginals, thus preserving a constraint of pairwise temporal consistency that the mean-field assumption lacks. Due to this greater representation of temporal constraints (the approximate posteriors at each time step being no longer assumed to be independent), the Bethe free energy has the potential to be significantly more accurate than the standard mean-field variational free energy (and is, in fact, exact for factor graphs without cycles such as the standard nonhierarchical POMDP model). In this letter, we focus entirely on the standard mean-field variational free energy used in the vast majority of active inference publications, and thus the Bethe free energy is out of scope for this article. However, exploring the nature of any intrinsic terms that might arise from the Bethe free energy is an interesting avenue for future work. Although primarily focused on the Bethe free energy, Schwöbel et al. (2018) also introduced a “predicted free energy” functional. This functional is equivalent to the FEF as we have defined it here, and so has a complexity instead of an information gain term, leading to minimizing the prior-posterior divergence.

Finally, Biehl et al. (2018) suggested that if the EFE is not mandated by the free-energy principle, which we have argued for in this letter, then in theory any standard intrinsic measure, such as empowerment, could be used as an objective. We believe that exploring the effect of these other potential loss functions could be a area of great interest for future work.

Acknowledgments

B.M. is supported by an EPSRC-funded PhD studentship. A.T. is funded by a PhD studentship from the Dr. Mortimer and Theresa Sackler Foundation and the School of Engineering and Informatics at the University of Sussex. C.L.B. is supported by BBRSC grant BB/P022197/1. A.T. is grateful to the Dr. Mortimer and Theresa Sackler Foundation, which supports the Sackler Centre for Consciousness Science.

References

- Attias, H. (2003). Planning by probabilistic inference. In *Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics*.
- Baldi, P., & Itti, L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Networks*, 23(5), 649–666.

- Baltieri, M., & Buckley, C. L. (2017). An active inference implementation of phototaxis. In *Proceedings of the Artificial Life Conference* (pp. 36–43). Berlin: Springer-Verlag.
- Baltieri, M., & Buckley, C. L. (2018). A probabilistic interpretation of PID controllers using active inference. In *From Animals to Animats: Proceedings of the International Conference on Simulation of Adaptive Behavior* (pp. 15–26). Cambridge, MA: MIT Press.
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76(4), 695–711.
- Beal, M. J. (1998). *Variational algorithms for approximate Bayesian inference*. PhD diss., University of London.
- Biehl, M., Guckelsberger, C., Salge, C., Smith, S. C., & Polani, D. (2018). Expanding the active inference landscape: more intrinsic motivations in the perception-action loop. *Frontiers in Neurobotics*, 12, 45.
- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518), 859–877.
- 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.
- Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., & Efros, A. A. (2018). *Large-scale study of curiosity-driven learning*. arXiv:1808.04355.
- Calvo, P., & Friston, K. (2017). Predicting green: Really radical (plant) predictive processing. *Journal of the Royal Society Interface*, 14(131), 20170096.
- Çatal, O., Verbelen, T., Nauta, J., De Boom, C., & Dhoedt, B. (2020). *Learning perception and planning with deep active inference*. arXiv:2001.11841.
- Cullen, M., Davey, B., Friston, K. J., & Moran, R. J. (2018). Active inference in OpenAI Gym: A paradigm for computational investigations into psychiatric illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(9), 809–818.
- Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., & Friston, K. (2020). *Active inference on discrete state-spaces: A synthesis*. arXiv:2001.07203.
- Deneve, S. (2005). Bayesian inference in spiking neurons. In L. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems*, 17 (pp. 353–360). Cambridge, MA: MIT Press.
- Doya, K., Ishii, S., Pouget, A., & Rao, R. P. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. Cambridge, MA: MIT Press.
- FitzGerald, T. H., Schwartenbeck, P., Moutoussis, M., Dolan, R. J., & Friston, K. (2015). Active inference, evidence accumulation, and the urn task. *Neural Computation*, 27(2), 306–328.
- Fox, C. W., & Roberts, S. J. (2012). A tutorial on variational Bayesian inference. *Artificial Intelligence Review*, 38(2), 85–95.
- Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, 16(9), 1325–1352.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836.

- Friston, K. (2008a). Hierarchical models in the brain. *PLOS Computational Biology*, 4(11).
- Friston, K. J. (2008b). Variational filtering. *NeuroImage*, 41(3), 747–766.
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.
- Friston, K. (2011). What is optimal about motor control? *Neuron*, 72(3), 488–498.
- Friston, K. (2019). *A free energy principle for a particular physics*. arXiv:1906.10184.
- Friston, K., & Ao, P. (2012). Free energy, value, and attractors. *Computational and Mathematical Methods in Medicine*, 2012, 937860.
- Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLOS One*, 4(7).
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience and Biobehavioral Reviews*, 68, 862–879.
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, 29(1), 1–49.
- Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology–Paris*, 100(1–3), 70–87.
- Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., & Ondobaka, S. (2017). Active inference, curiosity and insight. *Neural Computation*, 29(10), 2633–2683.
- Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, 104(1–2), 137–160.
- Friston, K. J., Parr, T., & de Vries, B. (2017). The graphical brain: Belief propagation and active inference. *Network Neuroscience*, 1(4), 381–414.
- Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, 6(4), 187–214.
- Friston, K. J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. *Neuroscience and Biobehavioral Reviews*, 90, 486–501.
- Friston, K. J., Trujillo-Barreto, N., & Daunizeau, J. (2008). DEM: A variational treatment of dynamic systems. *NeuroImage*, 41(3), 849–885.
- Houthoofd, R., Chen, X., Duan, Y., Schulman, J., De Turck, F., & Abbeel, P. (2016). Variational information maximizing exploration. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems*, 29. Red Hook, NY: Curran.
- Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, 49(10), 1295–1306.
- Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668), 20140169.
- Kappen, H. J. (2005). Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11), P11011.
- Kappen, H. J. (2007). An introduction to stochastic control theory, path integrals and reinforcement learning. In *AIP Conference Proceedings* (Vol. 887, pp. 149–181). College Park, MD: American Institute of Physics.

- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12), 712–719.
- Levine, S. (2018). *Reinforcement learning and control as probabilistic inference: Tutorial and review*. arXiv:1805.00909.
- Millidge, B. (2019a). *Combining active inference and hierarchical predictive coding: A tutorial introduction and case study*. <https://psyarxiv.com/kf6wc>
- Millidge, B. (2019b). *Implementing predictive processing and active inference: Preliminary steps and results*. <https://psyarxiv.com/4hb58/>
- Millidge, B. (2020). Deep active inference as variational policy gradients. *Journal of Mathematical Psychology*, 96, 102348.
- Millidge, B., Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). *On the relationship between active inference and control as inference*. arXiv:2006.12964.
- Mirza, M. B., Adams, R. A., Mathys, C. D., & Friston, K. J. (2016). Scene construction, visual foraging, and active inference. *Frontiers in Computational Neuroscience*, 10, 56.
- Mirza, M. B., Adams, R. A., Parr, T., & Friston, K. (2019). Impulsivity and active inference. *Journal of Cognitive Neuroscience*, 31(2), 202–220.
- Mohamed, S., & Rezende, D. J. (2015). Variational information maximisation for intrinsically motivated reinforcement learning. In C. Cortea, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems*, 28 (pp. 2125–2133). Red Hook, NY: Curran.
- Ostwald, D., Spitzer, B., Guggenmos, M., Schmidt, T. T., Kiebel, S. J., & Blankenburg, F. (2012). Evidence for neural encoding of Bayesian surprise in human somatosensation. *NeuroImage*, 62(1), 177–188.
- Oudeyer, P.-Y., & Kaplan, F. (2009). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, 1, 6.
- Parr, T. (2019). *The computational neurology of active vision*. PhD diss., University College London.
- Parr, T., Da Costa, L., & Friston, K. (2020). Markov blankets, information geometry and stochastic thermodynamics. *Philosophical Transactions of the Royal Society A*, 378(2164), 20190159.
- Parr, T., & Friston, K. J. (2017a). The active construction of the visual world. *Neuropsychologia*, 104, 92–101.
- Parr, T., & Friston, K. J. (2017b). Uncertainty, epistemics and active inference. *Journal of the Royal Society Interface*, 14(136), 20170376.
- Parr, T., & Friston, K. J. (2018a). Active inference and the anatomy of oculomotion. *Neuropsychologia*, 111, 334–343.
- Parr, T., & Friston, K. J. (2018b). The computational anatomy of visual neglect. *Cerebral Cortex*, 28(2), 777–790.
- Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological Cybernetics*, 113(5–6), 495–513.
- Parr, T., Markovic, D., Kiebel, S. J., & Friston, K. J. (2019). Neuronal message passing using mean-field, Bethe, and marginal approximations. *Scientific Reports*, 9(1), 1–18.
- Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 16–17). Piscataway, NJ: IEEE.

- Pezzulo, G., Cartoni, E., Rigoli, F., Pio-Lopez, L., & Friston, K. (2016). Active inference, epistemic value, and vicarious trial and error. *Learning and Memory*, 23(7), 322–338.
- Rawlik, K. C. (2013). *On probabilistic inference approaches to stochastic optimal control*. PhD diss., University of Edinburgh.
- Rawlik, K., Toussaint, M., & Vijayakumar, S. (2013). On stochastic optimal control and reinforcement learning by approximate inference. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press.
- Schwartenbeck, P., FitzGerald, T., Dolan, R., & Friston, K. (2013). Exploration, novelty, surprise, and free energy minimization. *Frontiers in Psychology*, 4, 710.
- Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *Elife*, 8, e41703.
- Schwöbel, S., Kiebel, S., & Marković, D. (2018). Active inference, belief propagation, and the Bethe approximation. *Neural Computation*, 30(9), 2530–2567.
- Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in Psychology*, 7, 1792.
- Spratling, M. W. (2008). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience*, 2, 4.
- Still, S., & Precup, D. (2012). An information-theoretic approach to curiosity-driven reinforcement learning. *Theory in Biosciences*, 131(3), 139–148.
- Sun, Y., Gomez, F., & Schmidhuber, J. (2011). Planning to be surprised: Optimal Bayesian exploration in dynamic environments. In *Proceedings of the International Conference on Artificial General Intelligence* (pp. 41–51). Berlin: Springer-Verlag.
- Theodorou, E., Buchli, J., & Schaal, S. (2010). A generalized path integral control approach to reinforcement learning. *Journal of Machine Learning Research*, 11, 3137–3181.
- Theodorou, E. A., & Todorov, E. (2012). Relative entropy and free energy dualities: Connections to path integral and KL control. In *Proceedings of the IEEE 51st Conference on Decision and Control* (pp. 1466–1473). Piscataway, NJ: IEEE.
- Toussaint, M. (2009). Probabilistic inference as a model of planned behavior. *KI*, 23(3), 23–29.
- Tschantz, A., Baltieri, M., Seth, A., & Buckley, C. L. (2019). *Scaling active inference*. arXiv:1911.10601.
- Tschantz, A., Millidge, B., Seth, A. K., & Buckley, C. L. (2020). *Reinforcement learning through active inference*. arXiv:2002.12636.
- Tschantz, A., Seth, A. K., & Buckley, C. L. (2019). *Learning action-oriented models through active inference*. bioRxiv:764969.
- Ueltzhöffer, K. (2018). Deep active inference. *Biological Cybernetics*, 112(6), 547–573.
- van de Laar, T. W., & de Vries, B. (2019). Simulating active inference processes by message passing. *Frontiers in Robotics and AI*, 6(20).
- Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2), Q1–305.

- Williams, G., Aldrich, A., & Theodorou, E. A. (2017). Model predictive path integral control: From theory to parallel computation. *Journal of Guidance, Control, and Dynamics*, *40*(2), 344–357.
- Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2001). Generalized belief propagation. In T. Leen, T. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems*, (pp. 689–695). Cambridge, MA: MIT Press.
- Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, *51*(7), 2282–2312.

Received July 5, 2020; accepted September 22, 2020.