ACTIVE INFERENCE ON DISCRETE STATE-SPACES − A SYNTHESIS

Lancelot Da Costa[1,2,*], **Thomas Parr**[2], **Noor Sajid**[2], **Sebastijan Veselic**[2], **Victorita Neacsu**[2], **Karl Friston**[2]

[1]Department of Mathematics, Imperial College London

[2]Wellcome Centre for Human Neuroimaging, University College London

January 22, 2020

ABSTRACT

Active inference is a normative principle underwriting perception, action, planning, decision-making and learning in biological or artificial agents. From its inception, its associated process theory has grown to incorporate complex generative models, enabling simulation of a wide range of complex behaviours. Due to successive developments in active inference, it is often difficult to see how its underlying principle relates to process theories and practical implementation. In this paper, we try to bridge this gap by providing a complete mathematical synthesis of active inference on discrete state-space models. This technical summary provides an overview of the theory, derives neuronal dynamics from first principles and relates this dynamics to biological processes. Furthermore, this paper provides a fundamental building block needed to understand active inference for mixed generative models; allowing continuous sensations to inform discrete representations. This paper may be used as follows: to guide research towards outstanding challenges, a practical guide on how to implement active inference to simulate experimental behaviour, or a pointer towards various in-silico neurophysiological responses that may be used to make empirical predictions.

Contents

---

*Author correspondence: l.da-costa@imperial.ac.uk

1 Introduction

Active inference is a normative principle underlying perception, action, planning, decision-making and learning in biological or artificial agents. It postulates that these processes may all be seen as optimising two complementary objective functions; namely, a variational free energy, which measures the fit between an internal model and (past) sensory observations, and an expected free energy, which scores possible (future) courses of action in relation to prior preferences. Active inference has been employed to simulate a wide range of complex behaviours, including planning and navigation [1], reading [2], curiosity and abstract rule learning [3], saccadic eye movements [4], visual foraging [5,6], visual neglect [7], hallucinations [8], niche construction [9,10], social conformity [11], impulsivity [12], image recognition [13] and the mountain car problem [14–16]. The key idea that underwrites these simulations is that creatures use an internal forward (generative) model to predict their sensory input, which they use to infer the causes of these data.

Early formulations of active inference employed generative models expressed in continuous space and time (for an introduction see [17], for a review see [18]), with behaviour modelled as a continuously evolving random dynamical system. However, we know that some processes in the brain conform better to discrete, hierarchical, representations, compared to continuous representations (e.g., visual working memory [19,20], state estimation via place cells [21,22], language, etc). Reflecting this, many of the paradigms studied in neuroscience are naturally framed as discrete state-space problems. Decision-making tasks are a prime candidate for this, as they often entail a series of discrete alternatives that an agent needs to choose among (e.g., multi-arm bandit tasks [23–25], multi-step decision tasks [26]). This explains why – in active inference – agent behaviour is often modelled using a discrete state-space formulation: the particular applications of which are summarised in Table 1. More recently, mixed generative models [27] – combining discrete and continuous states – have been used to model behaviour involving discrete and continuous representations (e.g., decision-making and movement [28], pharmacologically induced changes in eye-movement control [29] or reading; involving continuous visual sampling informing inferences about discrete semantics [27]).

Table 1: Applications of active inference (discrete state-space).

Application	Description	References
Decision-making under uncertainty	Initial formulation of active inference on partially observable Markov decision processes.	[30]
Optimal control	Application of KL or risk sensitive control in an engineering benchmark – the mountain car problem.	[14, 16]
Evidence accumulation	Illustrating the role of evidence accumulation in decision-making through an urns task.	[31, 32]
Psychopathology	Simulation of addictive choice behaviour.	[33]
Dopamine	The precision of beliefs about policies provides a plausible description of dopaminergic discharges.	[34, 35]
Functional magnetic resonance imaging	Empirical prediction and validation of dopaminergic discharges.	[36]
Maximal utility theory	Evidence in favour of surprise minimization as opposed to utility maximisation in human decision-making.	[37]
Social cognition	Examining the effect of prior preferences on interpersonal inference.	[38]
Exploration-exploitation dilemma	Casting behaviour as expected free energy minimising accounts for epistemic and pragmatic choices.	[39]
Habit learning and action selection	Formulating learning as an inferential process and action selection as Bayesian model averaging.	[40, 41]
Scene construction and anatomy of time	Mean-field approximation for multi-factorial hidden states, enabling high dimensional representations of the environment.	[5, 42]
Electrophysiological responses	Synthesising various in-silico neurophysiological responses via a gradient descent on free energy. E.g., place-cell activity, mismatch negativity, phase-precession, theta sequences, theta-gamma coupling and dopaminergic discharges.	[43]
Structure learning, curiosity and insight	Simulation of artificial curiosity and abstract rule learning. Structure learning via Bayesian model reduction.	[3]
Hierarchical temporal representations	Generalisation to hierarchical generative models with deep temporal structure and simulation of reading.	[2, 44]
Computational neuropsychology	Simulation of visual neglect, hallucinations, and prefrontal syndromes under alternative pathological priors.	[7, 45–48]

Neuromodulation	Use of precision parameters to manipulate exploration during saccadic searches; associating uncertainty with cholinergic and noradrenergic systems.	[6, 29, 49, 50]
Decisions to movements	Mixed generative models combining discrete and continuous states to implement decisions through movement.	[27, 28]
Planning, navigation and niche construction	Agent induced changes in environment (generative process); decomposition of goals into subgoals.	[1, 9, 10]
Atari games	Active inference compares favourably to reinforcement learning in the game of Doom.	[51]
Machine learning	Scaling active inference to more complex machine learning problems.	[52]

Due to the pace of recent theoretical advances in active inference, it is often difficult to retain a comprehensive overview of its process theory and practical implementation. In this paper, we hope to provide a comprehensive (mathematical) synthesis of active inference on discrete state-space models. This technical summary provides an overview of the theory, derives the associated (neuronal) dynamics from first principles and relates these to known biological processes. Furthermore, this paper and [18] provide the building blocks necessary to understand active inference on mixed generative models. This paper can be read as a practical guide on how to implement active inference for simulating experimental behaviour, or a pointer towards various in-silico neuro- and electro- physiological responses that can be tested empirically.

This paper is structured as follows. Section 2 is a high-level overview of active inference. The following sections elucidate the formulation by deriving the entire process theory from first principles; incorporating perception, planning and decision-making. This formalises the action-perception cycle: 1) an agent is presented with a stimulus, 2) it infers its latent causes, 3) plans into the future and 4) realises its preferred course of action; and repeat. This enactive cycle allows us to explore the dynamics of synaptic plasticity, which mediate learning of the contingencies of the world at slower timescales. We conclude in section 9 with an overview of structure learning in active inference.

2   Active inference

To survive in a changing environment, biological (and artificial) agents must maintain their sensations within a certain hospitable range (i.e., maintaining homeostasis through allostasis). In brief, active inference proposes that agents achieve this by optimising two complementary objective functions, a variational free energy and an expected free energy. In short, the former measures the fit between an internal (generative) model of its sensations and sensory observations, while the latter scores each possible course of action in terms of its ability to reach the range of "preferred" states of being.

Our first premise is that agents represent the world through an internal model. Through minimisation of variational free energy, this model becomes a good model of the environment. In other words, this probabilistic model and the probabilistic beliefs[2] that it encodes are continuously updated to mirror the environment and its dynamics. Such a world model is considered to be generative; in that it is able to generate predictions about sensations (e.g., during planning or dreaming), given beliefs about future states of being. If an agent senses a heat source (e.g., another agent) via some temperature receptors, the sensation of warmth represents an observed outcome and the temperature of the heat source a hidden state; minimisation of variational free energy then ensures that beliefs about hidden states closely match the true temperature. Formally, the generative model is a joint probability distribution over possible hidden states and sensory consequences – that specifies how the former cause the latter – and minimisation of variational free energy enables to "invert" the model; i.e., determine the most likely hidden states given sensations. The variational free energy is the negative evidence lower bound that is optimised in variational Bayes in machine learning [53, 54]. Technically – by minimising variational free energy – agents perform approximate Bayesian inference [55, 56], which enables them to infer the causes of their sensations (e.g., perception). This is the point of contact between active inference and the Bayesian brain [57–59]. Crucially, agents may incorporate an optimism bias [60, 61] in their model; thereby scoring certain "preferred" sensations as more likely. This lends a higher plausibility to those courses of action

---

[2]By beliefs we mean Bayesian beliefs, i.e., probability distributions over a variable of interest (e.g., current position). Beliefs are therefore used in the sense of Bayesian belief updating or belief propagation – as opposed to propositional or folk psychology beliefs.

that realise these sensations. In other words, a preference is simply something an agent (believes it) is likely to work towards.

To maintain homeostasis, and ensure survival, agents must minimise surprise[3]. Since the generative model scores preferred outcomes as more likely, minimising surprise corresponds to maximising model evidence[4]). In active inference, this is assured by the aforementioned processes; indeed, the variational free energy turns out to be an upper bound on surprise and minimising expected free energy ensures preferred outcomes are realised, thereby avoiding surprise on average.

Active inference can thus be framed as the minimisation of surprise [62–65] by perception and action. In discrete state models – of the sort discussed here – this means agents select from different possible courses of action (i.e., policies) in order to realise their preferences and thus minimise the surprise that they expect to encounter in the future. This enables a Bayesian formulation of the perception-action cycle [66]: agents perceive the world by minimising variational free energy, ensuring their model is consistent with past observations, and act by minimising expected free energy, to make future sensations consistent with their model. This account of behaviour can be concisely framed as self-evidencing [67].

In contrast to other normative models of behaviour, active inference is a 'first principle' account, which is grounded in statistical physics [68, 69]. Active inference describes the dynamics of systems that persist (i.e., do not dissipate) during some timescale of interest, and that can be statistically segregated from their environment – conditions which are satisfied by biological systems. Mathematically, the first condition means that the system is at *non-equilibrium steady-state* (NESS). This implies the existence of a steady-state probability density to which the system self-organises and returns to after perturbation (i.e., the agent's preferences). The statistical segregation condition is the presence of a Markov blanket (c.f., Figure 1) [70, 71]: a set of variables through which states internal and external to the system interact (e.g., the skin is a Markov blanket for the human body). Under these assumptions it can be shown that the states internal to the system parameterise Bayesian beliefs about external states and can be cast a process of variational free energy minimisation. This coincides with existing approaches to approximate inference [53, 72–74]. Furthermore, it can be shown that the most likely courses of action taken by those systems are those which minimise expected free energy – a quantity that subsumes many existing constructs in science and engineering (see section 7).

By subscribing to the above assumptions, it is possible to describe the behaviour of viable living systems as performing active inference. The remaining challenge is to determine the computational and physiological processes that implement to active inference. This paper aims to summarise a possible answers to this question, by reviewing the technical details of a process theory for active inference on discrete state-space generative models, first presented in [43]. Note that it is important to distinguish between active inference as a principle (presented above) from active inference as a process theory. The former is a consequence of fundamental assumptions about living systems, while the latter is a hypothesis concerning the computational and biological processes in the brain that might implement active inference. The ensuing process theories theory can then be used to predict plausible neuronal dynamics and electrophysiological responses that are elicited experimentally.

3   Discrete state-space generative models

The generative model [53] expresses how the agent represents the world. This is a joint probability distribution over sensory data and the hidden (or latent) causes of these data. The sorts of discrete state-space generative models used in active inference are specifically suited to represent discrete time series and decision-making tasks. These can be expressed as variants of partially observable Markov decision processes (POMDPs; [75]): from simple Markov decision processes [76–78] to generalisations in the form of deep probabilistic (hierarchical) models [2, 79, 80]. For clarity, the process theory is derived for the simplest model that facilitates understanding of subsequent generalisations; namely, a POMDP where the agent holds beliefs about the probability of the initial state (specified as D), the transition probabilities from one state to the next (defined as matrix B) and the probability of states given outcomes (i.e., the likelihood matrix A); see Figure 2.

---

[3]In information theory, the surprise (a.k.a., surprisal) associated with an outcome under a generative model is given by $-\log p(o)$. This specifies the extent to which an observation is unusual and surprises the agent – but this does not mean that the agent consciously experiences surprise. In information theory this kind of surprise is known as self-information.

[4]In Bayesian statistics, the model evidence (often referred to as marginal likelihood) associated with a generative model is $p(o)$ – the probability of observed outcomes according to the model (sometimes this is written as $p(o|m)$, explicitly conditioning upon a model). The model evidence scores the goodness of the model as an explanation of data that are sampled, by rewarding accuracy and penalising complexity, which avoids overfitting.

Table 2: Glossary of terms and notation.

Notation	Meaning	Type
S	Set of all possible (hidden) states.	Finite set of cardinality $m > 0$.
s_τ	(Hidden) state at time τ. In computations, if s_τ evaluates to the ith possible state, then interpret it as the ith unit vector in $\mathbb{R}^m$.	Random variable over S.
$s_{1:t}$	Sequence of hidden states $s_1, ..., s_t$.	Random variables over S^t.
O	Set of all possible outcomes.	Finite set of cardinality $n > 0$.
o_τ	Outcome at time τ. In computations, if o_τ evaluates to the jth possible outcome, then interpret it as the jth unit vector in $\mathbb{R}^n$.	Random variable over O.
$o_{1:t}$	Sequence of outcomes $o_1, ..., o_t$	Random variables over O^t.
T	Number of timesteps in a trial of observation epochs under the generative model.	Positive integer.
U	Set of all possible actions.	Finite set.
Π	Set of all allowable policies; i.e., sequences of actions.	Finite subset of U^T.
π	Policy.	Random variable over Π, or element of Π depending on context.
Q	Approximate posterior distribution.	Probability distribution over the latent variables of the generative model.
F, F_π	Variational free energy and variational free energy conditioned upon a policy.	Functionals of Q.
G	Expected free energy.	Function defined on Π.
Cat	Categorical distribution; probability distribution over a finite set assigning strictly positive probabilities.	Probability distribution over a finite set of cardinality k with parameter space $\{x \in \mathbb{R}^k \mid x_i > 0, \sum_i x_i = 1\}$
Dir	Dirichlet distribution (conjugate prior of the categorical distribution); probability distribution over the parameter space of the categorical distribution, parameterised by a vector of positive reals.	Probability distribution over $\{x \in \mathbb{R}^k \mid x_i > 0, \sum_i x_i = 1\}$, itself parameterised by an element of $(\mathbb{R}_{>0})^k$.
$X_{\bullet i}, X_{ki}$	ith column and (k,i)th element of matrix X.	Matrix indexing convention.
$\cdot, \otimes, \odot, {}^\odot$	Respectively inner product, Kronecker product, element-wise product and element-wise power. Following existing active inference literature, we adopt the convention $X \cdot Y := X^T Y$ for matrices.	Operation on vectors and matrices.
A	Likelihood matrix. The probability of the state-outcome pair o_τ, s_τ is given by $o_\tau \cdot A s_\tau$.	Random variable over the subset of $M_{n \times m}(\mathbb{R})$ with columns in $\{x \in \mathbb{R}^k \mid x_i > 0, \sum_i x_i = 1\}$.
B	Matrix of transition probabilities from one state to the next state given action $\pi_{\tau-1}$. The probability of possible state s_τ, given $s_{\tau-1}$ and action $\pi_{\tau-1}$ is $s_\tau \cdot B_{\pi_{\tau-1}} s_{\tau-1}$.	Matrix in $M_{m \times m}(\mathbb{R})$ with columns in $\{x \in \mathbb{R}^k \mid x_i > 0, \sum_i x_i = 1\}$.
$a, \mathbf{a}$	Parameters of prior and approximate posterior beliefs about A.	Matrices in $M_{n \times m}(\mathbb{R}_{>0})$.

6

$a_0, \mathbf{a}_0$	Matrices of the same size as $a, \mathbf{a}$, with homogeneous columns; any of its ith column elements are denoted by $a_{i0}, \mathbf{a}_{i0}$ and defined by $a_{i0} = \sum_{ji} a_{ji}, \mathbf{a}_{i0} = \sum_{ji} \mathbf{a}_{ji}.$	Matrices in $M_{n \times m}(\mathbb{R}_{>0})$.
$\log, \Gamma, \psi$	Natural logarithm, gamma function and digamma function. By convention these functions are taken component-wise on vectors and matrices.	Functions.
$\mathbb{E}_{P(X)}[f(X)]$	Expectation of random variable $f(X)$ under a probability density $P(X)$, taken component-wise if $f(X)$ is a matrix. $\mathbb{E}_{P(X)}[f(X)] := \int f(X) P(X)\, dX$	Real-valued operator on random variables.
$\mathbf{A}$	$\mathbf{A} := \mathbb{E}_{Q(A)}[A] = \mathbf{a} \odot \mathbf{a}_0^{\odot(-1)}$	Matrix in $M_{n \times m}(\mathbb{R}_{>0})$.
$\mathbf{logA}$	$\mathbf{logA} := \mathbb{E}_{Q(A)}[\log A] = \psi(\mathbf{a}) - \psi(\mathbf{a}_0)$. Note that $\mathbf{logA} \neq \log \mathbf{A}$!	Matrix in $M_{n \times m}(\mathbb{R})$.
σ	Softmax function or normalised exponential. $\sigma(x)_k = \frac{e^{x_k}}{\sum_i e^{x_i}}$	Function $\mathbb{R}^k \to \{x \in \mathbb{R}^k \| x_i > 0, \sum_i x_i = 1\}$
$\mathrm{H}[P]$	Shannon entropy of a probability distribution P. Explicitly, $\mathrm{H}[P] = \mathbb{E}_{P(x)}[-\log P(x)]$	Functional over probability distributions.

As mentioned above, a substantial body of work justifies describing certain neuronal representations with discrete state-space generative models (e.g., [19, 20, 86]). Furthermore, it has been long known that – at the level of neuronal populations – computations occur periodically (i.e., in distinct and sometimes nested oscillatory bands). Similarly, there is evidence for sequential computation in a number of processes (e.g., attention [87–89], visual perception [90, 91]) and at different levels of the neuronal hierarchy [2, 92], in line with ideas from hierarchical predictive processing [93, 94]. This accommodates the fact that visual saccadic sampling of observations occurs at a frequency of approximately 4Hz [28]. The relatively slow presentation of a discrete sequence of observations enables inferences to be performed in peristimulus time by (much) faster neuronal dynamics.

Active inference, implicitly, accounts for fast and slow neuronal dynamics. At each time-step the agent observes an outcome, from which it infers the past, present and future (hidden) states through perception. This underwrites a plan into the future, by evaluating (the expected free energy of) possible policies. The inferred (best) policies specify the most likely action, which is executed. At a slower timescale, parameters encoding the contingencies of the world (e.g., A), are inferred. This is referred to as learning. Even more slowly, the structure of the generative model is updated to better account for available observations – this is called structure learning. The following sections elucidate these aspects of the active inference process theory.

This paper will be largely concerned with deriving and interpreting the inferential dynamics that agents might implement using the generative model in Figure 2. We leave the discussion of more complex models to Appendix A, since the derivations are analogous in those cases.

4 Variational Bayesian inference

4.1 Free energy and model evidence

Variational Bayesian inference rests upon minimisation of a quantity called (variational) free energy, which measures the improbability (i.e., the surprise) of sensory observations, under a generative model. Simultaneously, variational free energy minimisation is a statistical inference technique that enables the approximation of the posterior distribution in Bayes rule. In machine learning, this is known as variational Bayes [53, 72–74]. Active inference agents minimise variational free energy, enabling concomitant maximisation of their model evidence and inference of the latent variables of their generative model. In the following, we consider a particular time point to be given $t \in \{1, ..., T\}$, whence the agent has observed a sequence of outcomes $o_{1:t}$. The posterior about the latent causes of sensory data is given by Bayes rule:
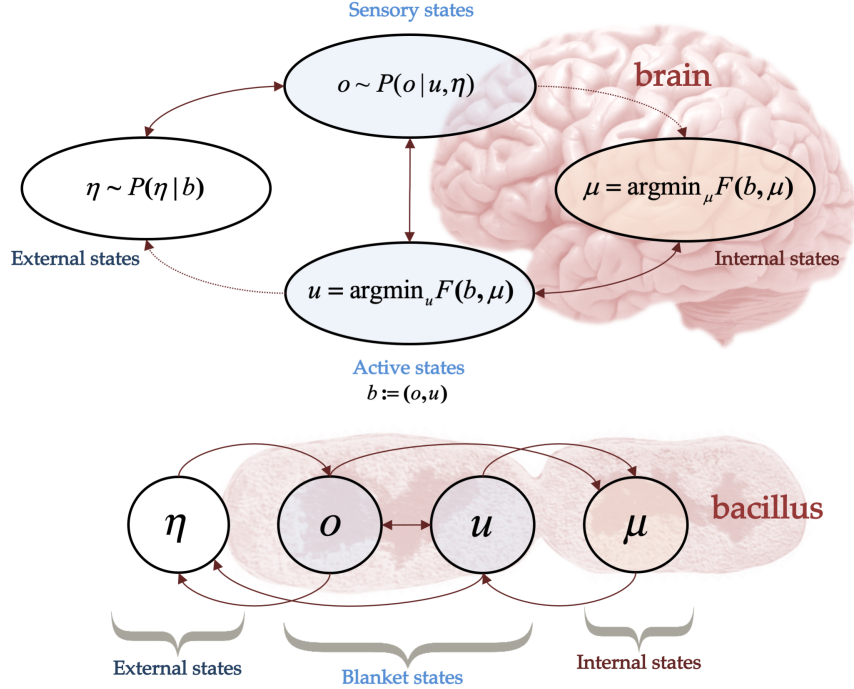
Figure 1: **Markov blankets in active inference.** This figure illustrates the Markov blanket assumption of active inference. A Markov blanket is a set of variables through which states internal and external to the system interact. Specifically, the system must be such that we can partition it into a Bayesian network of internal states μ, external states η, sensory states o and active states u, (μ, o and u are often referred together as *particular states*) with probabilistic (causal) links in the directions specified by the arrows. All interactions between internal and external states are therefore mediated by the blanket states b. The sensory states represent the sensory information that the body receives from the environment and the active states express how the body influences the environment. This blanket assumption is quite generic, in that it can be reasonably assumed for a brain as well as elementary organisms. For example, when considering a bacillus, the sensory states become the cell membrane and the active states comprise the actin filaments of the cytoskeleton. Under the Markov blanket assumption – together with the assumption that the system persists over time (i.e., possesses a non-equilibrium steady state) – a generalised synchrony appears, such that the dynamics of the internal states can be cast as performing inference over the external states (and vice-versa) via a minimisation of variational free energy [68, 69]. This coincides with existing approaches to inference; i.e., variational Bayes [53, 72–74]. This can be viewed as the internal states mirroring external states, via sensory states (e.g., perception), and external states mirroring internal states via active states (e.g., a generalised form of self-assembly, autopoiesis or niche construction). Furthermore, under these assumptions the most likely courses of actions can be shown to minimise expected free energy. Note that external states beyond the system should not be confused with the hidden states of the agent's generative model (which model external states). In fact, the internal states are exactly the parameters (i.e., sufficient statistics) encoding beliefs about hidden states and other latent variables, which model external states in a process of variational free energy minimisation. Hidden and external states may or may not be isomorphic. In other words, an agent uses its internal states to represent hidden states that may or may not exist in the external world.

2a

Generative model

$$P(o_{1:T}, s_{1:T}, A, \pi) = P(\pi)P(A)P(s_1)\prod_{\tau=2}^{T} P(s_\tau \mid s_{\tau-1}, \pi)\prod_{\tau=1}^{T} P(o_\tau \mid s_\tau, A)$$

Factors

1 $P(\pi) = Cat(E)$ — Prior beliefs about policies

$P(s_\tau)/P(o_\tau) = Cat(C)$ — Preferences over states/outcomes

2 $P(o_\tau \mid s_\tau, A) = Cat(As_\tau)$ — Likelihood mapping from hidden states to outcomes

3 $P(s_\tau \mid s_{\tau-1}, \pi) = Cat(B_{\pi_{\tau-1}} s_{\tau-1})$ — Transition probabilities from previous state to next state contingent on policies

4 $P(A) = \prod_{i=1}^{m} P(A_{\cdot i}), \quad P(A_{\cdot i}) = Dir(a_{\cdot i})$ — Beliefs about likelihood mapping

5 $P(s_1) = Cat(D)$ — Beliefs about the initial hidden state

2c

Approximate posterior

$$Q(s_{1:T}, A \mid \pi) = Q(A)Q(\pi)\prod_{\tau=1}^{T} Q(s_\tau \mid \pi)$$ — Mean-field approximation

Factors

$Q(s_\tau \mid \pi) = Cat(\mathbf{s}_{\pi\tau})$ — Approximate posterior over hidden states contingent on policies

$Q(A) = \prod_{i=1}^{m} Q(A_{\cdot i}), \quad Q(A_{\cdot i}) = Dir(\mathbf{a}_{\cdot i})$ — Approximate posterior over hidden state-outcome mapping

$Q(\pi) = Cat(\boldsymbol{\pi})$ — Approximate posterior over policies
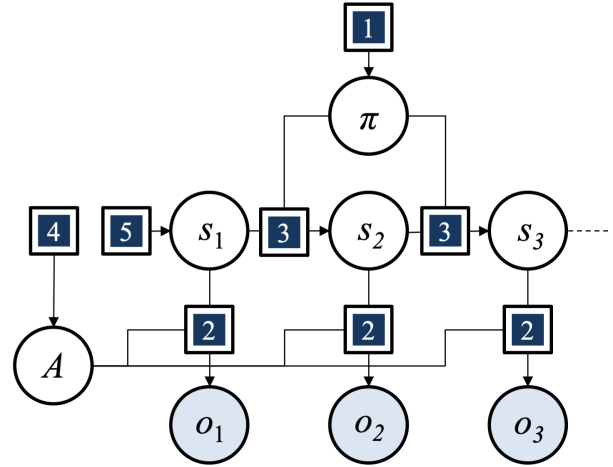
2b

Figure 2: **Example of a discrete state-space generative model.** Panel 2a, specifies the form of the generative model, which is how the agent represents the world. The generative model is a joint probability distribution over (hidden) states, outcomes and other variables that cause outcomes. In this representation, states unfold in time causing an observation at each time-step. The likelihood matrix A encodes the probabilities of state-outcome pairs. The policy π specifies which action to perform at each time-step. Note that the agent's preferences may be specified either in terms of states or outcomes. It is important to distinguish between states (resp. outcomes) that are random variables, and the possible values that they can take in S (resp. in O), which we refer to as possible states (resp. possible outcomes). Note that this type of representation comprises a finite number of timesteps, actions, policies, states, outcomes, possible states and possible outcomes. In Panel 2b, the generative model is displayed as a probabilistic graphical model [53, 71, 74, 81] expressed in factor graph form [82]. The variables in circles are random variables, while squares represent factors, whose specific form are given in Panel 2a. The arrows represent causal relationships (i.e., conditional probability distributions). The variables highlighted in grey can be observed by the agent, while the remaining variables are inferred through approximate Bayesian inference (see section 4) and called hidden or latent variables. Active inference agents perform inference by optimising the parameters of an approximate posterior distribution (see section 4). Panel 2c specifies how this approximate posterior factorises under a particular mean-field approximation [83], although other factorisations may be used [84, 85]. A glossary of terms used in this figure is available in Table 2. The mathematical yoga of generative models is heavily dependent on Markov blankets. The Markov blanket of a random variable in a probabilistic graphical model are those variables that share a common factor. Crucially, a variable conditioned upon its Markov blanket is conditionally independent of all other variables. We will use this property extensively (and implicitly) in the text.

$$P(s_{1:T}, A, \pi | o_{1:t}) = \frac{P(o_{1:t}|s_{1:T}, A, \pi)P(s_{1:T}, A, \pi)}{P(o_{1:t})} \tag{1}$$

Computing the posterior distribution requires computing the model evidence $P(o_{1:t}) = \sum_{\pi \in \Pi} \sum_{s_{1:T} \in S^T} \int P(o_{1:t}, s_{1:T}, A, \pi) \, dA$, which is intractable for complex generative models embodied by biological and artificial systems [92] – a well-known problem in Bayesian statistics. An alternative to computing the exact posterior distribution is to optimise an approximate posterior distribution over latent causes $Q(s_{1:T}, A, \pi)$, by minimising the Kullback-Leibler (KL) divergence [95] (D_{KL}) – a non-negative measure of discrepancy between probability distributions. We can use the definition of the KL divergence and Bayes rule to arrive at the variational free energy F, which is a functional of approximate posterior beliefs:

$$
\begin{aligned}
0 &\leq D_{KL}[Q(s_{1:T}, A, \pi) || P(s_{1:T}, A, \pi | o_{1:t})] \\
&= \mathbb{E}_{Q(s_{1:T}, A, \pi)}[\log Q(s_{1:T}, A, \pi) - \log P(s_{1:T}, A, \pi | o_{1:t})] \\
&= \mathbb{E}_{Q(s_{1:T}, A, \pi)}[\log Q(s_{1:T}, A, \pi) - \log P(o_{1:t}, s_{1:T}, A, \pi) + \log P(o_{1:t})] \\
&= \underbrace{\mathbb{E}_{Q(s_{1:T}, A, \pi)}[\log Q(s_{1:T}, A, \pi) - \log P(o_{1:t}, s_{1:T}, A, \pi)]}_{F[Q(s_{1:T}, A, \pi)]} + \log P(o_{1:t}) \\
\Rightarrow &- \log P(o_{1:t}) \leq F[Q(s_{1:T}, A, \pi)]
\end{aligned}
\tag{2}
$$

From (2), one can see that by varying Q to minimise the variational free energy enables us to approximate the true posterior, while simultaneously ensuring that surprise remains low. This means that variational free energy minimising agents, simultaneously, infer the latent causes of their observations and maximise the evidence for their generative model. To aid intuition, the variational free energy can be rearranged into complexity and accuracy:

$$F[Q(s_{1:T}, A, \pi)] = \underbrace{D_{KL}[Q(s_{1:T}, A, \pi) || P(s_{1:T}, A, \pi)]}_{\text{Complexity}} - \underbrace{\mathbb{E}_{Q(s_{1:T}, A, \pi)}[\log P(o_{1:t}|s_{1:T}, A, \pi)]}_{\text{Accuracy}} \tag{3}$$

The first term of (3) can be regarded as complexity: a simple explanation for observable data Q, which makes few assumptions over and above the prior (i.e., with KL divergence close to zero), is a good explanation. In other words, a good explanation is an accurate account of some data that requires minimal movement for updating of prior to posterior beliefs (c.f., Occam's principle). The second term is accuracy; namely, the probability of the data given posterior beliefs about model parameters Q. In other words, how well the generative model fits the observed data. The idea that neural representations weigh complexity against accuracy underwrites the imperative to find the most accurate explanation for sensory observations that is minimally complex, which has been leveraged by things like Horace Barlow's principle of minimum redundancy [96] and subsequently supported empirically [97–100]. Figure 3 illustrates the various implications of minimising free energy.

4.2 On the family of approximate posteriors

The goal is now to minimise variational free energy with respect to Q. To obtain a tractable expression for the variational free energy, we need to assume a certain simplifying factorisation of the approximate posterior. There are many possible forms [84, 117, 118] (e.g., mean-field, marginal, Bethe), each of which trades off the quality of the inferences with the complexity of the computations involved. For the purpose of this paper we use a particular (structured) mean-field approximation (c.f., Figure 2):

$$Q(s_{1:T}, A, \pi) = Q(A)Q(\pi) \prod_{\tau=1}^{T} Q(s_\tau | \pi) \tag{4}$$

This choice is driven by didactic purposes and since this factorisation has been used extensively in the active inference literature (e.g., [2, 27, 43]). However, the most recent software implementation of active inference (i.e., `spm_MDP_VB_X.m`) employs a marginal approximation [84, 119], which retains the simplicity and biological interpretation of the neuronal dynamics afforded by the mean-field approximation, while approximating the more accurate inferences of the Bethe approximation. For these reasons, the marginal free energy currently stands as the most biologically plausible.

Self-evidencing
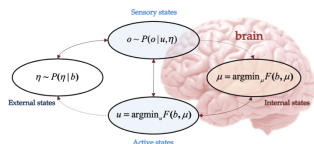
$$\mu = \arg\min_{\mu} F(b,\mu) \quad \text{Perception}$$

$$u = \arg\min_{u} F(b,\mu) \quad \text{Action}$$

$-F(b,\mu) \leq \ln P(b\|m) =$ **Value**	Reinforcement learning Optimal control theory Expected utility theory
$F(b,\mu) \geq -\ln P(b\|m) =$ **Surprise**	Infomax principle Minimum redundancy Free-energy principle
$\mathbb{E}\big[F(b,\mu)\big] \geq \mathrm{H}[P(b\|m)] =$ **Entropy**	Self-organization Synergetics Homeostasis
$P(b\|m) =$ **Model evidence**	Bayesian brain Evidence accumulation Predictive coding

Pavlov

Barlow

Haken

Helmholtz

Figure 3: **Markov blankets and self-evidencing.** This schematic illustrates the various interpretations of minimising variational free energy. Recall that the existence of a Markov blanket implies a certain lack of influences among internal, blanket and external states. These independencies have an important consequence; internal and active states are the only states that are not influenced by external states, which means their dynamics (i.e., perception and action) are a function of, and only of, particular states (i.e., internal, sensory and active states); here, the variational (free energy) bound on surprise. This surprise has a number of interesting interpretations. Given it is the negative log probability of finding a particle or creature in a particular state, minimising surprise corresponds to maximising the value of a particle's state. This interpretation is licensed by the fact that the states with a high probability are, by definition, attracting states. On this view, one can then spin-off an interpretation in terms of reinforcement learning [76], optimal control theory [101] and, in economics, expected utility theory [102]. Indeed, any scheme predicated on the optimisation of some objective function can now be cast in terms of minimising surprise – in terms of perception and action (i.e., the dynamics of internal and active states) – by specifying these optimal values to be the agent's preferences. The minimisation of surprise (i.e., self-information) leads to a series of influential accounts of neuronal dynamics; including the principle of maximum mutual information [103, 104], the principles of minimum redundancy and maximum efficiency [105] and the free energy principle [65]. Crucially, the average or expected surprise (over time or particular states of being) corresponds to entropy. This means that action and perception look as if they are minimising entropy. This leads us to theories of self-organisation, such as synergetics in physics [106–108] or homeostasis in physiology [109–111]. Finally, the probability of any blanket states given a Markov blanket (m) is, on a statistical view, model evidence [112, 113]. This means that all the above formulations are internally consistent with things like the Bayesian brain hypothesis, evidence accumulation and predictive coding; most of which inherit from Helmholtz's motion of unconscious inference [114], later unpacked in terms of perception as hypothesis testing in 20th century psychology [115] and machine learning [116].

4.3 Computing the variational free energy

The next sections focus on producing biologically plausible neuronal dynamics that perform perception and learning based on variational free energy minimisation. To enable this, we first compute variational the free energy, using the factorisations of the generative model and approximate posterior (c.f., Figure 2):

$$
\begin{aligned}
F[Q(s_{1:T}, A, \pi)] &= \mathbb{E}_{Q(s_{1:T}, A, \pi)}[\log Q(s_{1:T}, A, \pi) - \log P(o_{1:t}, s_{1:T}, A, \pi)] \\
&= \mathbb{E}_{Q(s_{1:T}, A, \pi)}\Big[\log Q(A) + \log Q(\pi) + \sum_{\tau=1}^{T} \log Q(s_\tau|\pi) \\
&\quad - \log P(A) - \log P(\pi) - \log P(s_1) - \sum_{\tau=2}^{T} \log P(s_\tau|s_{\tau-1}, \pi) - \sum_{\tau=1}^{t} \log P(o_\tau|s_\tau, A)\Big] \\
&= D_{KL}[Q(A)||P(A)] + D_{KL}[Q(\pi)||P(\pi)] + \mathbb{E}_{Q(\pi)}[F_\pi[Q(s_{1:T}|\pi)]]
\end{aligned}
\tag{5}
$$

where

$$
\begin{aligned}
F_\pi[Q(s_{1:T}|\pi)] &:= \sum_{\tau=1}^{T} \mathbb{E}_{Q(s_\tau|\pi)}[\log Q(s_\tau|\pi)] - \sum_{\tau=1}^{t} \mathbb{E}_{Q(s_\tau|\pi)Q(A)}[\log P(o_\tau|s_\tau, A)] \\
&\quad - \mathbb{E}_{Q(s_1|\pi)}[\log P(s_1)] - \sum_{\tau=2}^{T} \mathbb{E}_{Q(s_\tau|\pi)Q(s_{\tau-1}|\pi)}[\log P(s_\tau|s_{\tau-1}, \pi)]
\end{aligned}
\tag{6}
$$

is the variational free energy conditioned upon a policy. This is the same quantity that we would have obtained by omitting A and conditioning all probability distributions in the numerators of (1) by π. In the next section, we will see how perception can be framed in terms of variational free energy minimisation.

5 Perception

In active inference, perception is equated with state estimation [43] (e.g., inferring the temperature from the sensation of warmth), consistent with the idea that perceptions are hypotheses [115]. To infer the (past, present and future) states of the environment, an agent must minimise the variational free energy with respect to $Q(s_{1:T}|\pi)$ for each policy π. This provides the agent's inference over hidden states, contingent upon pursuing a given policy. Since the only part of the free energy that depends on $Q(s_{1:T}|\pi)$ is F_π, the agent must simply minimise F_π. Substituting $Q(s_\tau|\pi)$ by their sufficient statistics (i.e., parameters $\mathbf{s}_{\pi\tau}$), F_π becomes a function of those parameters. This enables us to rewrite (6), conveniently in matrix form:

$$
\begin{aligned}
F_\pi(\mathbf{s}_{\pi 1}, ..., \mathbf{s}_{\pi T}) &= \sum_{\tau=1}^{T} \mathbf{s}_{\pi\tau} \cdot \log \mathbf{s}_{\pi\tau} - \sum_{\tau=1}^{t} o_\tau \cdot \mathbf{logA}\mathbf{s}_{\pi\tau} \\
&\quad - \mathbf{s}_{\pi 1} \log D - \sum_{\tau=2}^{T} \mathbf{s}_{\pi\tau} \cdot \log(B_{\pi_{\tau-1}})\mathbf{s}_{\pi\tau-1}
\end{aligned}
\tag{7}
$$

This enables to compute the variational free energy gradients [120]:

$$
\nabla_{\mathbf{s}_{\pi\tau}} F_\pi(\mathbf{s}_{\pi 1}, ..., \mathbf{s}_{\pi T}) = \vec{1} + \log \mathbf{s}_{\pi\tau} - \begin{cases} o_\tau \cdot \mathbf{logA} + \mathbf{s}_{\pi\tau+1} \cdot \log(B_{\pi_\tau}) + \log D & \text{if } \tau = 1 \\ o_\tau \cdot \mathbf{logA} + \mathbf{s}_{\pi\tau+1} \cdot \log(B_{\pi_\tau}) + \log(B_{\pi_{\tau-1}})\mathbf{s}_{\pi\tau-1} & \text{if } 1 < \tau \leq t \\ \mathbf{s}_{\pi\tau+1} \cdot \log(B_{\pi_\tau}) + \log(B_{\pi_{\tau-1}})\mathbf{s}_{\pi\tau-1} & \text{if } \tau > t \end{cases}
\tag{8}
$$

The neuronal dynamics are given by a gradient descent on free energy [43], with state-estimation expressed as a softmax function of accumulated (negative) free energy gradients. The constant term $\vec{1}$ is generally omitted since the softmax function removes it anyway. This enables us to equate the gradient with a prediction error.

$$
\begin{aligned}
\dot{v}(\mathbf{s}_{\pi 1}, ..., \mathbf{s}_{\pi T}) &= -\nabla_{\mathbf{s}_{\pi\tau}} F_\pi(\mathbf{s}_{\pi 1}, ..., \mathbf{s}_{\pi T}) \\
\mathbf{s}_{\pi\tau} &= \sigma(v)
\end{aligned}
\tag{9}
$$

The softmax function – a generalisation of the sigmoid to vector inputs – is a natural choice as the variational free energy gradient is a logarithm and the components of $\mathbf{s}_{\pi\tau}$ must sum to one.

5.1 Plausibility of neuronal dynamics

The temporal dynamics expressed in (9) unfold at a much faster timescale than the sampling of new observations (i.e., within timesteps) and correspond to fast neuronal processing in peristimulus time. This is consistent with behaviour-relevant computations at frequencies that are higher than the rate of visual sampling (e.g., working memory [121], visual stimulus perception in humans [90] and macaques [91]).

Furthermore, these dynamics are consistent with predictive processing [122, 123] – since active inference prescribes dynamics that minimise prediction error – although they generalise it to a wide range of generative models. Note that, while also a variational free energy gradient, this sort of prediction error is not the same as that given by predictive coding schemes (which rely upon a different sort of generative model) [17, 18, 124].

Just as neuronal dynamics involve translation from post-synaptic potentials to firing rates, (9) involves translating from a vector of real numbers (v), to a vector whose elements are bounded between zero and one ($\mathbf{s}_{\pi\tau}$); via the softmax function. As a result, it is natural to interpret the components of v as the average membrane potential of distinct neural populations, and $\mathbf{s}_{\pi\tau}$ as the average firing rate of those populations, which is bounded thanks to neuronal refractory periods. This is consistent with mean-field formulations of neural population dynamics, in that the average firing rate of a neuronal population follows a sigmoid function of the average membrane potential [125–127]. Using the fact that a softmax function is a generalisation of the sigmoid to vector inputs – here the average membrane potentials of coupled neuronal populations – it follows that their average firing follows a softmax function of their average potential. In this context, the softmax function may be interpreted as performing lateral inhibition, which can be thought of as leading to narrower tuning curves of individual neurons and thereby sharper inferences [128]. Importantly, this tells us that state-estimation can be performed in parallel by different neuronal populations, and a simple neuronal architecture is sufficient to implement these dynamics (see Figure 6 in [84]).

Lastly, interpreting the dynamics in this way has a degree of face validity, as it enables us to synthesise a wide-range of biologically plausible electrophysiological responses; including repetition suppression, mismatch negativity, violation responses, place-cell activity, phase precession, theta sequences, theta-gamma coupling, evidence accumulation, race-to-bound dynamics and transfer of dopamine responses [36, 43].

The neuronal dynamics for state estimation coincide with variational message passing [129, 130]: a widely used algorithm for approximate Bayesian inference. This is an important result, since it shows that variational message passing emerges under active inference using a particular mean-field approximation. If one were to use the Bethe approximation, the corresponding dynamics coincide with belief propagation [53, 82, 84, 85, 117], another widely used algorithm for approximate inference. This offers a formal connection between active inference and message passing interpretations of neuronal dynamics [27, 131, 132]. In the next section, we examine planning, decision-making and action selection.

6 Planning, decision-making and action selection

So far, we have focused on optimising beliefs about hidden states under a particular policy by minimising a variational free energy functional of an approximate posterior over hidden states, under each policy.

In this section, we explain how planning and decision-making arise as a minimisation of expected free energy – a function scoring the goodness of each possible future course of action. We briefly motivate how the expected free energy arises from first-principles. This allows us to frame decision-making and action-selection in terms of expected free energy minimisation. Finally, we conclude by discussing the computational cost of planning into the future.

6.1 Planning and decision-making

At the heart of active inference, is a description of agents that strive to attain a target distribution specifying the range of preferred states of being, given a sufficient amount of time. To work towards reaching these preferences, agents select policies $Q(\pi)$, such that their predicted states $Q(s_\tau, A)$ at some future time point $\tau > t$ (usually, the time horizon of a policy T) reach the preferred states $P(s_\tau, A)$, which are specified by the generative model. These considerations

allow us to show in Appendix B that the requisite approximate posterior over policies $Q(\pi)$ is a softmax function of the negative *expected free energy* G[5]:

$$
\begin{aligned}
Q(\pi) &= \sigma(-G(\pi)) \\
G(\pi) &= \underbrace{D_{KL}[Q(s_\tau, A|\pi)||P(s_\tau, A)]}_{Risk} - \underbrace{\mathbb{E}_{Q(s_\tau, A|\pi)P(o_\tau|s_\tau, A)}[\log P(o_\tau|s_\tau, A)]}_{Ambiguity}
\end{aligned}
\tag{10}
$$

This means that the most likely (i.e., best) policies minimise expected free energy. This ensures that future courses of action are exploitative (i.e., *risk* minimising) and explorative (i.e., *ambiguity* minimising). In particular, the expected free energy specifies the optimal balance between goal-seeking and itinerant novelty-seeking behaviour, given some prior preferences or goals. Note that the ambiguity term rests on an expectation over fictive (i.e., predicted) outcomes under beliefs about future states. This means that optimising beliefs about future states during perception is crucial to accurately predict future outcomes during planning. In summary, planning and decision-making respectively correspond to evaluating the expected free energy of different policies, which scores their goodness in relation to prior preferences and forming approximate posterior beliefs about policies.

6.2   Action selection, policy-independent state-estimation

Approximate posterior beliefs about policies allows to obtain the most plausible action as the most likely under all policies – this can be expressed as a Bayesian model average

$$
u_t = \arg\max_{u \in U} \left(\sum_{\pi \in \Pi} \delta_{u, \pi_t} Q(\pi) \right)
\tag{11}
$$

where δ is the Kronecker delta. In addition, we obtain a policy independent state-estimation at any time point $\tau \in \{1, ..., T\}$, as a Bayesian model average of approximate posterior beliefs about hidden states under policies:

$$
\begin{aligned}
Q(s_\tau) &= \sum_{\pi \in \Pi} Q(s_\tau|\pi)Q(\pi) \\
\iff \mathbf{s}_\tau &= \sum_{\pi \in \Pi} \mathbf{s}_{\pi\tau} Q(\pi)
\end{aligned}
\tag{12}
$$

Note that these Bayesian model averages may be implemented by neuromodulatory mechanisms [41].

6.3   Biological plausibility

Winner take-all architectures of decision-making are already commonplace in computational neuroscience (e.g., models of selective attention and recognition [133, 134], hierarchical models of vision [135]). This is nice, since the softmax function in (10) can be seen as providing a biologically plausible [125–127], smooth approximation to the maximum operation, which is known as soft winner take-all [136]. In fact, the generative model, presented in Figure 2, can be naturally extended such that the approximate posterior contains an (inverse) temperature parameter γ multiplying the expected free energy inside the softmax function (see Appendix A.2). This temperature parameter regulates how precisely the softmax approximates the maximum function, thus recovering winner take-all architectures for high parameter values (technically, this converts Bayesian model averaging into Bayesian model selection, where the policy corresponds to a model of what the agent is doing). This parameter, regulating precision of policy selection, has a clear biological interpretation in terms of confidence encoded in dopaminergic firing [34–36, 43]. Interestingly, Daw and colleagues [23] uncovered evidence in favour of a similar model employing a softmax function and temperature parameter in human decision-making.

---

[5]A more complete treatment may include priors over policies – usually denoted by E – and the evidence for a policy afforded by observed outcomes (usually denoted by $\mathbf{F}$). These additional terms supplement the expected free energy, leading to an approximate posterior of the form $\sigma(-\log E - \mathbf{F} - \mathbf{G})$ [2].

6.4 Pruning of policy trees

From a computational perspective, planning (i.e., computing the expected free energy) for each possible policy can be cost-prohibitive, due do the combinatorial explosion in the number of sequences of actions when looking deep into the future. There has been work in understanding how the brain finesses this problem [137], which suggests a simple answer: during mental planning, humans stop evaluating a policy as soon as they encounter a large loss (i.e., a high value of the expected free energy that renders the policy highly implausible). In active inference this corresponds to using an Occam window; that is, we stop evaluating the expected free energy of a policy if it becomes much higher than the best (smallest expected free energy) policy – and set its approximate posterior probability to an arbitrarily low value accordingly. This biologically plausible pruning strategy drastically reduces the number of policies one has to evaluate exhaustively.

Although effective and biologically plausible, the Occam window for pruning policy trees cannot deal with large policy spaces that ensue with deep policy trees and long temporal horizons. This means that pruning can only partially explain how biological organisms perform deep policy searches. Further research is needed to characterise the processes in which biological agents reduce large policy spaces to tractable subspaces. One explanation – for the remarkable capacity of biological agents to evaluate deep policy trees – rests on deep (hierarchical) generative models, in which policies operate at each level. These deep models enable long-term policies, modelling slow transitions among hidden states at higher levels in the hierarchy, to contextualise faster state transitions at subordinate levels (see Appendix A). The resulting (semi Markovian) process can then be specified in terms of a hierarchy of limited horizon policies that are nested over temporal scales; c.f., motor chunking [138–140].

6.5 Discussion of the action-perception cycle

Minimising variational and expected free energy are complementary and mutually beneficial processes. Minimisation of variational free energy ensures that the generative model is a good predictor of its environment; this allows the agent to accurately plan into the future by evaluating expected free energy, which in turn enables it to realise its preferences. In other words, minimisation of variational free energy is a vehicle for effective planning and reaching preferences via the expected free energy; in turn, reaching preferences minimises the expected surprise of future states of being.

In conclusion, we have seen how agents plan into the future and make decisions about the best possible course of action. This concludes our discussion of the action-perception cycle. In the next section, we examine expected free energy in greater detail. Then, we will see how active agents can learn the contingencies of the environment and the structure of their generative model at slower timescales.

7 Properties of the expected free energy

The expected free energy is a fundamental construct of interest. In this section, we unpack its main features and highlight its importance in relation to many existing theories in neurosciences and engineering.

The expected free energy of a policy can be unpacked in a number of ways. Perhaps the most intuitive is in terms of risk and ambiguity:

$$G(\pi) = \underbrace{D_{KL}[Q(s_\tau, A|\pi)||P(s_\tau, A)]}_{\text{Risk}} + \underbrace{\mathbb{E}_{Q(s_\tau, A|\pi)}[\mathrm{H}[P(o_\tau|s_\tau, A)]]}_{\text{Ambiguity}} \tag{13}$$

This means that policy selection minimises risk and ambiguity. Risk, in this setting, is simply the difference between predicted and prior beliefs about final states. In other words, policies will be deemed more likely if they bring about states that conform to prior preferences. In the optimal control literature, this part of expected free energy underwrites KL control [141, 142]. In economics, it leads to risk sensitive policies [143]. Ambiguity reflects the uncertainty about future outcomes, given hidden states. Minimising ambiguity therefore corresponds to choosing future states that generate unambiguous and informative outcomes (e.g., switching on a light in the dark).

We can express the expected free energy of a policy as a bound on information gain and expected log (model) evidence (a.k.a., Bayesian risk):

$$G(\pi) = \underbrace{\mathbb{E}_Q[D_{KL}[Q(s_\tau, A|o_\tau, \pi)||P(s_\tau, A|o_\tau)]]}_{\text{Expected evidence bound}} - \underbrace{\mathbb{E}_Q[\log P(o_\tau)]}_{\text{Expected log evidence}}$$
$$- \underbrace{\mathbb{E}_Q[D_{KL}[Q(s_\tau, A|o_\tau, \pi)||Q(s_\tau, A|\pi)]]}_{\text{Expected information gain}} \tag{14}$$
$$\geq - \underbrace{\mathbb{E}_Q[\log P(o_\tau)]}_{\text{Expected log evidence}} - \underbrace{\mathbb{E}_Q[D_{KL}[Q(s_\tau, A|o_\tau, \pi)||Q(s_\tau, A|\pi)]]}_{\text{Expected information gain}}$$

The first term in (14) is the expectation of log evidence under beliefs about future outcomes, while the second ensures that this expectation is maximally informed, when outcomes are encountered. Collectively, these two terms underwrite the resolution of uncertainty about hidden states (i.e., information gain) and outcomes (i.e., expected surprise) in relation to prior beliefs.

When the agent's preferences are expressed in terms of outcomes (c.f., Figure 2), it is useful to express risk in terms of outcomes, as opposed to hidden states. This is most useful when the generative model is not known or during structure learning, when the state-space evolves over time. In these cases, the risk over hidden states can be replaced risk over outcomes by assuming the KL divergence between the predicted and true posterior (under expected outcomes) is small:

$$\underbrace{D_{KL}[Q(s_\tau, A|\pi)||P(s_\tau, A)]}_{\text{Risk (states)}} = \underbrace{D_{KL}[Q(o_\tau|\pi)||P(o_\tau)]}_{\text{Risk (outcomes)}} + \underbrace{\mathbb{E}_{Q(o_\tau|\pi)}[D_{KL}[Q(s_\tau, A|o_\tau, \pi)||P(s_\tau, A|o_\tau)]]}_{\approx 0}$$
$$\approx \underbrace{D_{KL}[Q(o_\tau|\pi)||P(o_\tau)]}_{\text{Risk (outcomes)}} \tag{15}$$

This divergence constitutes an expected evidence bound that also appears if we express expected free energy in terms of intrinsic and extrinsic value:

$$G(\pi) = - \underbrace{\mathbb{E}_{Q(o_\tau|\pi)}[\log P(o_\tau)]}_{\text{Extrinsic value}} + \underbrace{\mathbb{E}_{Q(o_\tau|\pi)}[D_{KL}[Q(s_\tau, A|o_\tau, \pi)||P(s_\tau, A|o_\tau)]]}_{\text{Expected evidence bound}}$$
$$- \underbrace{\mathbb{E}_{Q(o_\tau|\pi)}[D_{KL}[Q(s_\tau|o_\tau, \pi)||Q(s_\tau|\pi)]]}_{\text{Intrinsic value (states) or salience}} - \underbrace{\mathbb{E}_{Q(o_\tau, s_\tau|\pi)}[D_{KL}[Q(A|o_\tau, s_\tau, \pi)||Q(A)]]}_{\text{Intrinsic value (parameters) or novelty}} \tag{16}$$

Extrinsic value is just the expected value of log evidence, which can be associated with reward and utility in behavioural psychology and economics, respectively [144–146]. In this setting, extrinsic value is the negative of Bayesian risk [147], when reward is log evidence. The intrinsic value of a policy is its epistemic value or affordance [39]. This is just the expected information gain afforded by a particular policy, which can be about hidden states (i.e., salience) or model parameters (i.e., novelty). It is this term that underwrites artificial curiosity [148].

Intrinsic value is also known as intrinsic motivation in neurorobotics [144, 149, 150], the value of information in economics [151], salience in the visual neurosciences and (rather confusingly) Bayesian surprise in the visual search literature [152–154]. In terms of information theory, intrinsic value is mathematically equivalent to the expected mutual information between hidden states in the future and their consequences – consistent with the principles of minimum redundancy or maximum efficiency [104, 105, 155]. Finally, from a statistical perspective, maximising intrinsic value (i.e., salience and novelty) corresponds to optimal Bayesian design [156] and machine learning derivatives, such as active learning [157]. On this view, active learning is driven by novelty; namely, the information gain afforded model parameters, given future states and their outcomes. Heuristically, this curiosity resolves uncertainty about "what would happen if I did that" [146]. Figure 4 illustrates the compass of expected free energy, in terms of its special cases; ranging from optimal Bayesian design through to Bayesian decision theory.

8 Learning

In active inference, learning concerns the dynamics of synaptic plasticity, which are thought to encode beliefs about the contingencies of the environment [43] (e.g., beliefs about B, in some settings, are thought to be encoded in recurrent
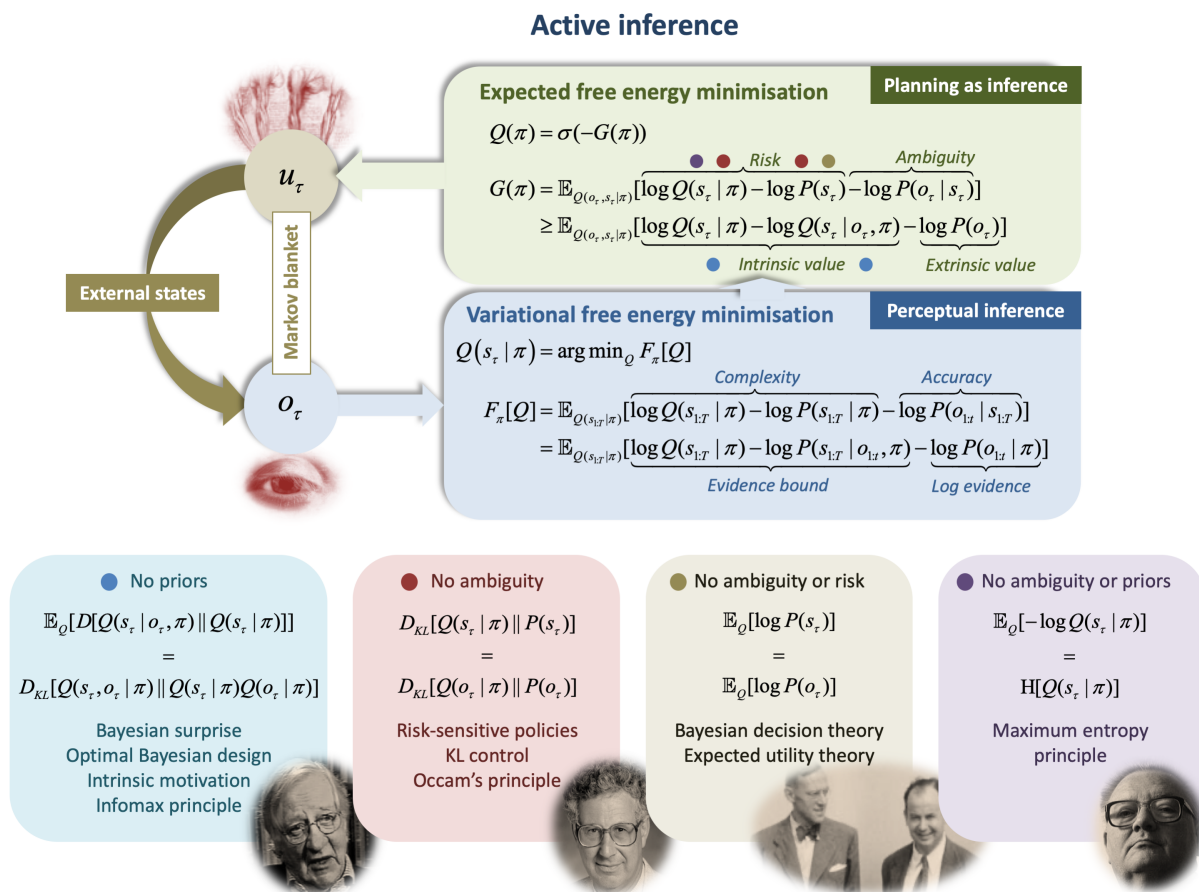
Active inference

Expected free energy minimisation   **Planning as inference**

$$Q(\pi) = \sigma(-G(\pi))$$

$$G(\pi) = \mathbb{E}_{Q(o_\tau, s_\tau | \pi)}[\underbrace{\log Q(s_\tau | \pi) - \log P(s_\tau)}_{Risk} - \underbrace{\log P(o_\tau | s_\tau)}_{Ambiguity}]$$

$$\geq \mathbb{E}_{Q(o_\tau, s_\tau | \pi)}[\underbrace{\log Q(s_\tau | \pi) - \log Q(s_\tau | o_\tau, \pi)}_{Intrinsic\ value} - \underbrace{\log P(o_\tau)}_{Extrinsic\ value}]$$

Variational free energy minimisation   **Perceptual inference**

$$Q(s_\tau | \pi) = \arg\min_Q F_\pi[Q]$$

$$F_\pi[Q] = \mathbb{E}_{Q(s_{1:T} | \pi)}[\underbrace{\log Q(s_{1:T} | \pi) - \log P(s_{1:T} | \pi)}_{Complexity} - \underbrace{\log P(o_{1:t} | s_{1:T})}_{Accuracy}]$$

$$= \mathbb{E}_{Q(s_{1:T} | \pi)}[\underbrace{\log Q(s_{1:T} | \pi) - \log P(s_{1:T} | o_{1:t}, \pi)}_{Evidence\ bound} - \underbrace{\log P(o_{1:t} | \pi)}_{Log\ evidence}]$$

External states · **Markov blanket** · u_τ · o_τ

No priors	No ambiguity	No ambiguity or risk	No ambiguity or priors
$\mathbb{E}_Q[D[Q(s_\tau \| o_\tau, \pi) \| Q(s_\tau \| \pi)]]$	$D_{KL}[Q(s_\tau \| \pi) \| P(s_\tau)]$	$\mathbb{E}_Q[\log P(s_\tau)]$	$\mathbb{E}_Q[-\log Q(s_\tau \| \pi)]$
$=$	$=$	$=$	$=$
$D_{KL}[Q(s_\tau, o_\tau \| \pi) \| Q(s_\tau \| \pi) Q(o_\tau \| \pi)]$	$D_{KL}[Q(o_\tau \| \pi) \| P(o_\tau)]$	$\mathbb{E}_Q[\log P(o_\tau)]$	$H[Q(s_\tau \| \pi)]$
Bayesian surprise Optimal Bayesian design Intrinsic motivation Infomax principle	Risk-sensitive policies KL control Occam's principle	Bayesian decision theory Expected utility theory	Maximum entropy principle

Figure 4: **Expected free energy.** This figure illustrates the various ways in which minimising expected free energy can be unpacked (omitting model parameters for clarity). The upper panel casts action and perception as the minimisation of variational and expected free energy, respectively. Crucially, active inference introduces beliefs over policies that enable a formal description of planning as inference [1, 158, 159]. In brief, posterior beliefs about hidden states of the world, under plausible policies, are optimised by minimising a variational (free energy) bound on log evidence. These beliefs are then used to evaluate the expected free energy of allowable policies, from which actions can be selected [43]. Crucially, expected free energy subsumes several special cases that predominate in the psychological, machine learning and economics literature. These special cases are disclosed when one removes particular sources of uncertainty from the implicit optimisation problem. For example, if we ignore prior preferences, then the expected free energy reduces to information gain [112, 156] or intrinsic motivation [144, 149, 150]. This is mathematically the same as expected Bayesian surprise and mutual information that underwrite salience in visual search [152, 154] and the organisation of our visual apparatus [103–105, 155]. If we now remove risk but reinstate prior preferences, one can effectively treat hidden and observed (sensory) states as isomorphic. This leads to risk sensitive policies in economics [143, 160] or KL control in engineering [142]. Here, minimising risk corresponds to aligning predicted outcomes to preferred outcomes. If we then remove ambiguity and relative risk of action (i.e., intrinsic value), we are left with extrinsic value or expected utility in economics [161] that underwrites reinforcement learning and behavioural psychology [76]. Bayesian formulations of maximising expected utility under uncertainty is also known as Bayesian decision theory [147]. Finally, if we just consider a completely unambiguous world with uninformative priors, expected free energy reduces to the negative entropy of posterior beliefs about the causes of data; in accord with the maximum entropy principle [162]. The expressions for variational and expected free energy correspond to those described in the main text (omitting model parameters for clarity). They are arranged to illustrate the relationship between complexity and accuracy, which become risk and ambiguity, when considering the consequences of action. This means that risk-sensitive policy selection minimises expected complexity or computational cost. The coloured dots above the terms in the equations correspond to the terms that constitute the special cases in the lower panels.

excitatory connections in the prefrontal cortex [46]). The fact that beliefs about matrices (e.g., A, B) may be encoded in synaptic weights conforms to connectionist models of brain function, as it offers a convenient way to compute probabilities, in the sense that the synaptic weights could be interpreted as performing matrix multiplication as in artificial neural networks, to predict; e.g., outcomes from beliefs about states, using the likelihood matrix A.

These synaptic dynamics (e.g., long-term potentiation and depression) evolve at a slower timescale than action and perception, which is consistent with the fact that such inferences need evidence accumulation over multiple state-outcome pairs. For simplicity, we will assume the only variable that is learned is A, but what follows generalises to more complex generative models (c.f., Appendix A.1. Learning A means that approximate posterior beliefs about A follow a gradient descent on variational free energy. Seeing the variational free energy (5) as a function of $\mathbf{a}$ (the sufficient statistic of $Q(A)$) we can write:

$$
\begin{aligned}
F(\mathbf{a}) &= D_{KL}[Q(A)||P(A)] - \sum_{\tau=1}^{t} \mathbb{E}_{Q(\pi)Q(s_\tau|\pi)Q(A)}[o_\tau \cdot \log(A)s_\tau] + \cdots \\
&= D_{KL}[Q(A)||P(A)] - \sum_{\tau=1}^{t} o_\tau \cdot \mathbf{logA}s_\tau + \cdots
\end{aligned}
\tag{17}
$$

Here, we ignore the terms in (5) that do not depend on $Q(A)$, as these will vanish when we take the gradient. The KL-divergence between Dirichlet distributions is [163, 164]:

$$
\begin{aligned}
D_{KL}[Q(A)||P(A)] &= \sum_{i=1}^{m} D_{KL}[Q(A_{\cdot i})||P(A_{\cdot i})] \\
&= \sum_{i=1}^{m} \left(\log \Gamma(\mathbf{a}_{0i}) - \sum_{k=1}^{n} \log \Gamma(\mathbf{a}_{ki}) - \log \Gamma(a_{0i}) + \sum_{k=1}^{n} \log \Gamma(a_{ki}) + (\mathbf{a}_{\bullet i} - a_{\bullet i}) \cdot (\mathbf{logA})_{\bullet i} \right) \\
&= \sum_{i=1}^{m} \left(\log \Gamma(\mathbf{a}_{0i}) - \sum_{k=1}^{n} \log \Gamma(\mathbf{a}_{ki}) - \log \Gamma(a_{0i}) + \sum_{k=1}^{n} \log \Gamma(a_{ki}) \right) + (\mathbf{a} - a) \cdot \mathbf{logA}
\end{aligned}
\tag{18}
$$

Incorporating (18) in (17), we can take the gradient of the variational free energy with respect to $\mathbf{logA}$:

$$
\nabla_{\mathbf{logA}} F(\mathbf{a}) = \mathbf{a} - a - \sum_{\tau=1}^{t} o_\tau \otimes \mathbf{s}_\tau
\tag{19}
$$

where $\otimes$ is the Kronecker product. This means that the dynamics of synaptic plasticity follow a descent on (19):

$$
\begin{aligned}
\dot{\rho}(\mathbf{a}) &= -\nabla_{\mathbf{logA}} F(\mathbf{a}) \\
&= -\mathbf{a} + a + \sum_{\tau=1}^{t} o_\tau \otimes \mathbf{s}_\tau
\end{aligned}
\tag{20}
$$

In computational terms, these are the dynamics for evidence accumulation of Dirichlet parameters at time t. Since synaptic plasticity dynamics occur at a much slower pace than perceptual inference, it is computationally much cheaper – in numerical simulations – to do a one-step belief update at the end of each trial of observation epochs. Explicitly, setting the free energy gradient to zero at the end of the trial gives the following update for Dirichlet parameters:

$$
\mathbf{a} = a + \sum_{\tau=1}^{T} o_\tau \otimes \mathbf{s}_\tau
\tag{21}
$$

After which, the prior beliefs $P(A)$ are updated to the approximate posterior beliefs $Q(A)$ for the subsequent trial. Note that this update scheme is formally identical to associative or Hebbian plasticity.

As one can see, the learning rule concerning accumulation of Dirichlet parameters (c.f., (21)) means that the agent becomes increasingly confident about its likelihood matrix by receiving new observations (since the matrix which is added onto a at each timestep is always positive). This is fine as long as the structure of the environment remains relatively constant. In the next section, we will see how Bayesian model reduction can revert this process, to enable the agent to adapt quickly to a changing environment. Table 3 summarises the belief updating entailed by active inference, and Figure 5 indicates where particular computations might be implemented in the brain.

Table 3: Summary of belief updating.

Process	Computation	Equations
Perception	$\mathbf{s}_{\pi\tau} = \sigma(v), \dot{v} = -\nabla_{\mathbf{s}_{\pi\tau}} F_\pi$	(8)
Planning	$G(\pi)$	(43), (44)
Decision-making	$Q(\pi) = \sigma(-G(\pi))$	(10)
Action selection	$u_t = \arg\max_{u \in U} \left(\sum_{\pi \in \Pi} \delta_{u,\pi_t} Q(\pi) \right)$	(11)
Policy-independent state-estimation	$\mathbf{s}_\tau = \sum_{\pi \in \Pi} \mathbf{s}_{\pi\tau} Q(\pi)$	(12)
Learning (end of trial)	$\mathbf{a} = a + \sum_{\tau=1}^{T} o_\tau \otimes \mathbf{s}_\tau$	(21)

9 Structure learning

In the previous sections, we have addressed how an agent performs inference over different variables at different timescales in a biologically plausible fashion, which we equated to perception, planning and decision-making. In this section, we consider the problem of learning the form or structure of the generative model.

The idea here is that agents are equipped (e.g., born) with an innate generative model that entails fundamental preferences (e.g., essential to survival), which are not updated. For instance, humans are born with prior preferences about their body temperature around $37°$C and cardiac frequency within a certain range. Mathematically, this means that the parameters of these innate prior distributions – encoding the agent's expectations as part of its generative model – have hyperpriors that are infinitely precise (e.g., a Dirac delta distribution) and thus cannot be updated in an experience dependent fashion. The agent's generative model then naturally evolves by minimising variational free energy to become a good model of the agent's environment but is still constrained by the survival preferences hardcoded within it. This process of learning the generative model (i.e., the variables and their functional dependencies) is called structure learning.

Structure learning in active inference is an active area of research. Active inference proposes that the agent's generative model evolves over time to maximise the evidence for its observations. However, a complete set of mechanisms that biological agents use to do so has not yet been laid out. Nevertheless, we use this section to summarise two complementary approaches; namely, Bayesian model reduction and Bayesian model expansion [3, 171–173] – that enable to simplify and complexify the model, respectively.

9.1 Bayesian model reduction

To explain the causes of their sensations, agents must compare different hypotheses about how their sensory data are generated – and retain the hypothesis or model that is the most valid in relation to their observations (i.e., has the greatest model evidence). In Bayesian statistics, these processes are called Bayesian model comparison and Bayesian model selection – these correspond to scoring the evidence for various generative models in relation to available data and selecting the one with the highest evidence [174, 175]. Bayesian model reduction (BMR) is a particular instance of structure learning, which formalises *post-hoc* hypothesis testing to simplify the generative model. This precludes redundant explanations of sensory data – and ensures the model generalises to new data. Technically, it involves estimating the evidence for simpler (reduced) priors over the latent causes and selecting the model with the highest evidence. This process of simplifying the generative model – by removing certain states or parameters – has a clear biological interpretation in terms of synaptic decay and switching off certain synaptic connections, which
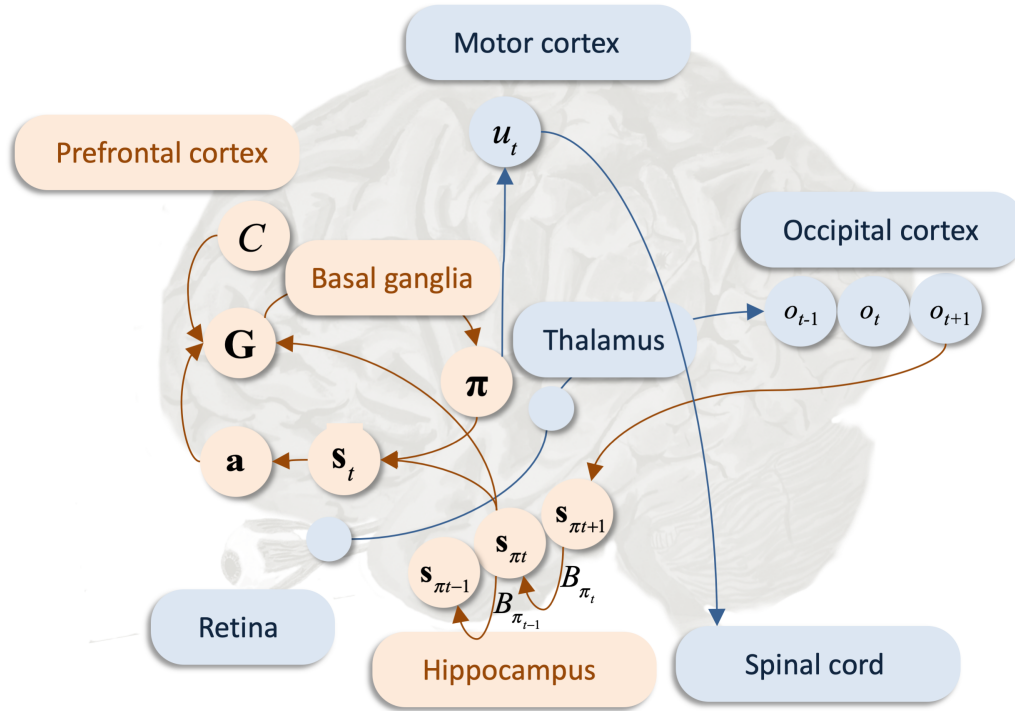
Figure 5: **Possible functional anatomy.** This figure summarises a possible (coarse-grained) functional anatomy that could implement belief updating in active inference. The arrows correspond to message passing between different neuronal populations. Here, a visual observation is sampled by the retina, aggregated in first-order sensory thalamic nuclei and processed in the occipital (visual) cortex. The green arrows correspond to message passing of sensory information. This signal is then propagated (via the ventral visual pathway) to inferior and medial temporal lobe structures such as the hippocampus; this allows the agent to go from observed outcomes to beliefs about their most likely causes in state-estimation (perception), which is performed locally. The variational free energy is computed in the striatum. The orange arrows encode message passing of beliefs. Preferences C are attributed to the dorsolateral prefrontal cortex – which is thought to encode representations over prolonged temporal scales [44] – consistent with the fact that these are likely to be encoded within higher cortical areas [3]. The expected free energy is computed in the medial prefrontal cortex [43] during planning, which leads to inferences about most plausible policies (decision-making) in the basal ganglia, consistent with the fact that the basal ganglia is thought to underwrite planning and decision-making [165–170]. The message concerning policy selection is sent to the motor cortex via thalamocortical loops. The most plausible action, which is selected in the motor cortex is passed on through the spinal cord to trigger a limb movement. Simultaneously, policy independent state-estimation is performed in the ventrolateral prefrontal cortex, which leads to synaptic plasticity dynamics in the prefrontal cortex, where the synaptic weights encode beliefs about A.

is reminiscent of the synaptic mechanisms of sleep (e.g., REM sleep [176, 177]), reflection and associated machine learning algorithms (e.g., wake-sleep [178]).

To keep things concise, let ν represent a hidden variable in the generative model that is optimised during learning (e.g. A), and $o = o_{1:t}$ a sequence of observations. The current model has a prior $P(\nu)$ and we would like to test whether a reduced prior (i.e., less complex) $\tilde{P}(\nu)$ can provide a more parsimonious explanation for the observed outcomes. Using Bayes rule, we have the following identities:

$$P(\nu)P(o|\nu) = P(\nu|o)P(o) \tag{22}$$

$$\tilde{P}(\nu)P(o|\nu) = \tilde{P}(\nu|o)\tilde{P}(o) \tag{23}$$

Where $P(o) = \int P(o|\nu)P(\nu)\, d\nu$ and $\tilde{P}(o) = \int P(o|\nu)\tilde{P}(\nu)$. Dividing (22) by (23) yields

$$\frac{P(\nu)}{\tilde{P}(\nu)} = \frac{P(\nu|o)P(o)}{\tilde{P}(\nu|o)\tilde{P}(o)} \tag{24}$$

We can then use (24) in order to obtain the following relations:

$$1 = \int \tilde{P}(\nu|o)\, d\nu = \frac{P(o)}{\tilde{P}(o)} \int \frac{\tilde{P}(\nu)P(\nu|o)}{P(\nu)}\, d\nu = \frac{P(o)}{\tilde{P}(o)} \mathbb{E}_{P(\nu|o)}\left[\frac{\tilde{P}(\nu)}{P(\nu)}\right] \tag{25}$$

$$\Rightarrow \log \tilde{P}(o) - \log P(o) = \log \mathbb{E}_{P(\nu|o)}\left[\frac{\tilde{P}(\nu)}{P(\nu)}\right] \tag{26}$$

We can approximate the posterior term in the expectation of (26) with the corresponding approximate posterior $Q(\nu)$, which simplifies the computation. This allows us to compare the evidence of the two models (reduced and full) and select the best. If the reduced model has more evidence, it implies the current model is too complex – and redundant parameters can be removed by adopting the new priors.

In conclusion, BMR allows for computationally efficient and biologically plausible hypothesis testing, to find simpler explanations for the data at hand. It has been used to emulate sleep and reflection in abstract rule learning [3], by simplifying the prior over A at the end of each trial – this has the additional benefit of preventing the agent from becoming overconfident.

9.2 Bayesian model expansion

Bayesian model expansion is complementary to Bayesian model reduction. It entails adopting a more complex generative model (by adding, e.g., more states); if, and only if the gain in accuracy in (3) is sufficient enough to outweigh the increase in complexity. This model expansion allows for generalisation and concept learning in active inference [172]. Note that additional states need not always lead to a more complex model. It is in principle possible to expand a model in such a way that complexity decreases, as many state estimates might be able to remain close to their priors in place of a small number of estimates moving a lot. This 'shared work' by many parameters could lead to a simpler model.

From a computational perspective, concept acquisition can be seen as a type of structure learning [179, 180] – that can be emulated through Bayesian model comparison. Recent work on concept learning in active inference [172], shows that a generative model equipped with extra (latent) hidden states can engage these 'unused' hidden states, when an agent is presented with novel stimuli during the learning process. Initially the corresponding likelihood mappings (i.e., the corresponding columns of A) are uninformative, but these are updated when the agent encounters new observations that cannot be accounted by its current knowledge (e.g., observing a cat when it has only been exposed to birds). This happens naturally, during the learning process, in an unsupervised way through free energy minimization. To allow for effective generalization, this approach can be combined with BMR; in which any new concept can be aggregated with similar concepts, and the associated likelihood mappings can be reset for further concept acquisition, in favour of a simpler model with higher model evidence. This approach can be further extended by updating the number of extra hidden states through a process of Bayesian model comparison.

10 Discussion

Due to the various recent theoretical advances in active inference, it is easy to lose sight of its underlying principle, process theory and practical implementation. We have tried to address this by rehearsing – in a clear and concise way – the assumptions underlying active inference as a principle, the technical details of the process theory for discrete state-space generative models and the biological interpretation of the accompanying neuronal dynamics. It is useful to clarify these results; as a first step to guide towards outstanding theoretical research challenges, a practical guide to implement active inference to simulate experimental behaviour and a pointer towards various predictions that may be tested empirically.

Active inference offers a degree of plausibility as a process theory of brain function. From a theoretical perspective its requisite neuronal dynamics correspond to known empirical phenomena and extend earlier theories like predictive coding [63, 122, 123]. Furthermore, the process theory is consistent with the underlying free energy principle, which biological systems are thought to abide by – namely, the avoidance of surprising states: this can be articulated formally based on fundamental assumptions about biological systems [68, 69]. Lastly, the process theory has a degree of face validity as its predicted electrophysiological responses closely resemble empirical measurements.

However, for a full endorsement of the process theory presented in this paper, rigorous empirical validation of the synthetic electrophysiological responses is needed. To pursue this, one would have to specify the generative model that a biological agent employs for a particular task. This can be done through Bayesian model comparison of alternative generative models with respect to empirical (choice) behaviour being measured (e.g., [181]). Once the appropriate generative model is formulated, evidence for a plausible but distinct implementations of active inference would need to be compared, which come from various possible approximations to the free energy [84, 85, 117], each of which yields different belief updates and simulated electrophysiological responses. Note that the marginal approximation to the free energy currently stands as the most biologically plausible [84]. From this, the explanatory power of active inference can be assessed in relation to empirical measurements and contrasted with other existing theories.

This means that the key challenge for active inference – and arguably data analysis in general – is finding the generative model that best explains observable data (i.e., evidence maximising). A solution to this problem would enable to find the generative model – entailed by an agent – by observing its behaviour. In turn, this would enable one to simulate its belief updating and behaviour accurately *in-silico*. It should be noted that these generative models can be specified manually for the purposes of reproducing simple behaviour (e.g., agents performing simple tasks needed for empirical validation discussed above). However, a generic solution to this problem is necessary to account for complex datasets; in particular, complex behavioural data from agents in a real environment. Moreover, a biologically plausible solution to this problem could correspond to a complete structure learning roadmap; accounting for how biological agents evolve their generative model to account for new observations. Evolution has solved this problem by selecting phenotypes with a good model of their sensory data, therefore, understanding the processes that have selected generative models that are fit for purpose for our environment might lead to important advances in structure learning and data analysis.

Discovering new generative models corresponding to complex behavioural data, will demand to extend the current process theory to these models, in order to provide testable predictions and reproduce the observed behaviour in-silico. Examples of generative models that do not fall within the current discrete state-space, continuous state-space [8, 18, 30, 182–185] or mixed [27–29] models – currently implemented in active inference – include Markov decision trees [74, 186] and Boltzmann machines [77, 187, 188].

One challenge that may arise, when scaling active inference to complex models with many degrees of freedom, will be the size of the policy trees in consideration. Although effective and biologically plausible, the current pruning strategy is unlikely to reduce the search space sufficiently to enable tractable inference in such cases. As noted above, the issue of scaling active inference may yield to the first principles of the variational free energy formulation. Specifically, generative models with a high evidence are minimally complex. This suggests that 'scaling up', in and of itself, is not the right strategy for reproducing more sophisticated or deep behaviour. A more principled approach would be to explore the right kind of factorisations necessary to explain structured behaviour. A key candidate here are deep temporal or diachronic generative models that have a separation of timescales. This form of factorisation (c.f., mean field approximation) replaces deep decision trees with shallow decision trees that are hierarchically composed.

To summarise, we argue that some important challenges for theoretical neuroscience include finding process theories of brain function that comply with active inference as a principle [68, 69]; namely, the avoidance of surprising events. The outstanding challenge is then to explore and fine grain such process theories, via Bayesian model comparison (e.g., using dynamic causal modelling [58, 189]) in relation to experimental data. From a structure learning and data analysis perspective, the main challenge is finding the generative model with the greatest evidence in relation to available data. This may be achieved by understanding the processes evolution has selected for creatures with a good model of their

environment. Finally, to scale active inference to behaviour with many degrees of freedom, one needs to understand how biological agents effectively search deep policy trees when planning into the future, when many possible policies may be entertained at separable timescales.

11   Conclusion

In conclusion, this paper aimed to summarise: the assumptions underlying active inference, the technical details underwriting its process theory, and how the associated neuronal dynamics relate to known biological processes. These processes underwrite action, perception, planning, decision-making, learning and structure learning; which we have illustrated under discrete state-space generative models. We have discussed some important outstanding challenges: from a broad perspective, the challenge for theoretical neuroscience is to develop increasingly fine-grained mechanistic models of brain function that comply with the core tenets of active inference [68, 69]. In regards to the process theory, key challenges relate to experimental validation, understanding how biological organisms evolve their generative model to account for new sensory observations and how they effectively search large policy spaces when planning into the future.

Software availability

The belief updating scheme described in this article is generic and can be implemented using standard routines (e.g., `spm_MDP_VB_X.m`). These routines are available as Matlab code in the SPM academic software: `http://www.fil.ion.ucl.ac.uk/spm/`. Examples of simulations using discrete state-space generative models can be found via a graphical user interface by typing DEM.

Acknowledgements

A   More complex generative models

In this Appendix, we briefly present cases of more complex discrete state-space generative models and explain how the belief updating can be extended to those cases.

A.1   Learning B and D

In this paper, we have only considered the case where A is learned, while beliefs about B (i.e., transition probabilities from one state to the next) and D (i.e., beliefs about the initial state) remained fixed. In general, B and D can also be learnt over time. This calls upon a new (extended) expression for the generative model with priors over B and D:

$$P(o_{1:T}, s_{1:T}, A, B, D, \pi) = P(\pi)P(A)P(B)P(D)P(s_1|D) \prod_{\tau=2}^{T} P(s_\tau|s_{\tau-1}, B, \pi) \prod_{\tau=1}^{T} P(o_\tau|s_\tau, A)$$

$$P(B) = \prod_{u \in U} \prod_{i=1}^{m} P((B_u)_{\bullet i}) \quad P((B_u)_{\bullet i}) = Dir((b_u)_{\bullet i})$$

$$P(D) = Dir(d)$$

(27)

Here, $(B_u)_{\bullet i}$ and $(b_u)_{\bullet i}$ denote the ith columns of the matrix B_u encoding the transition probabilities from one state to the next state and its corresponding Dirichlet parameter b_u. Furthermore, one needs to define the corresponding approximate posteriors that will be used for learning:

$$Q(B) = \prod_{u \in U} \prod_{i=1}^{m} Q((B_u)_{\bullet i}) \quad Q((B_u)_{\bullet i}) = Dir((\mathbf{b}_u)_{\bullet i})$$

$$Q(D) = Dir(d)$$

(28)

The variational free energy, after having observed $o_{1:t}$, is computed analogously as in equation (5). The process of finding the belief dynamics is then akin to section 8 – we rehearse it in the following: selecting only those terms in the variational free energy, which depend on B and D yields:

$$F[Q(B,D)] = D_{KL}[Q(B)||P(B)] + D_{KL}[Q(D)||P(D)] - \mathbb{E}_{Q(\pi)Q(s_1|\pi)Q(D)}[s_1 \cdot \log D]$$

$$- \sum_{\tau=2}^{t} \mathbb{E}_{Q(\pi)Q(s_\tau,s_{\tau-1}|\pi)Q(B)}[s_\tau \cdot \log B_{\pi_\tau} s_{\tau-1}] + \cdots$$

$$= D_{KL}[Q(B)||P(B)] + D_{KL}[Q(D)||P(D)] - \mathbf{s}_1 \cdot \mathbf{logD} - \sum_{\tau=2}^{t} \mathbb{E}_{Q(\pi)}[\mathbf{s}_{\pi\tau} \cdot \mathbf{logB}_{\pi_\tau} \mathbf{s}_{\pi_{\tau-1}}] + \cdots$$

(29)

Using the form of the KL divergence for Dirichlet distributions (18) and taking the gradients yields

$$\nabla_{\mathbf{logB}_u} F(\mathbf{b}_u) = \mathbf{b}_u - b_u - \sum_{\tau=2}^{t} \sum_{\pi \in \Pi} \delta_{u,\pi_t} Q(\pi)(\mathbf{s}_{\pi\tau} \otimes \mathbf{s}_{\pi\tau-1})$$

(30)

$$\nabla_{\mathbf{logD}} F(\mathbf{d}) = \mathbf{d} - d - \mathbf{s}_1$$

(31)

where $\otimes$ denotes the Kronecker product. Finally, it is possible to specify neuronal plasticity dynamics following a descent on (30), (31), which correspond to biological dynamics. Alternatively, we have belief update rules implemented once after each trial of observation epochs in in-silico agents:

$$\mathbf{b}_u = b_u + \sum_{\tau=2}^{t} \sum_{\pi \in \Pi} \delta_{u,\pi_\tau} Q(\pi)(\mathbf{s}_{\pi\tau} \otimes \mathbf{s}_{\pi\tau-1})$$

(32)

$$\mathbf{d} = d + \mathbf{s}_1$$

(33)

A.2 Complexifying the prior over policies

In this paper, we have considered a simple prior approximate posterior over policies; namely, $\sigma(-G(\pi))$. This can be extended to $\sigma(-\gamma G(\pi))$, where γ is an (inverse) temperature parameter that denotes the confidence in selecting a particular policy. This extension is quite natural in the sense that γ can be interpreted as the postsynaptic response to dopaminergic input [34, 35]. This correspondence is supported by empirical evidence [36] and enables one to simulate biologically plausible dopaminergic discharges (c.f., Appendix E [43]). Anatomically, this parameter may be encoded within the substantia nigra, in nigrostriatal dopamine projection neurons [36], which maps well with our proposed functional anatomy (c.f., Figure 5), since the substantia nigra is connected with the striatum. We refer the reader to [43] for a discussion of the associated belief updating scheme.

A.3 Multiple state and outcome modalities

In general, one does not only need one hidden state and outcome factor to represent the environment, but many. Intuitively, this happens in the human brain as we integrate sensory stimuli from our five (or more) distinct senses. Mathematically, we can express this via different streams of hidden states (usually referred to as hidden factors) that evolve independently of one another that interact to generate outcomes at each time step; e.g., see Figure 9 [74] for a graphical representation of a multi-factorial hidden Markov model. This means that A becomes a multi-dimensional tensor that integrates information about the different hidden factors to cause outcomes. The belief updating is analogous in this case, contingent upon the fact that one assumes a mean-field factorisation of the approximate posterior on the different hidden state factors (see, e.g., [5, 42]). This means that the beliefs about states may be processed in a manner analogous to Figure 5, invoking a greater number of neural populations.

A.4 Deep temporal models

A deep temporal model is a generative model with many layers that are nested hierarchically and act at different timescales. These were first introduced within active inference in [3]. One can picture them graphically as a POMDP (c.f., Figure 2) at the higher level where each outcome is replaced by a POMDP at the lower level, and so forth.

There is a useful metaphor for understanding the concept underlying deep temporal models: each layer of the model corresponds to the hand of a clock. In a two-layer hierarchical model, a ticking (resp. rotation) of the faster hand corresponds to a time step (resp. trial of observation epochs) at the lower level. At the end of each trial at the lower level, the slower hand ticks once, which corresponds to a time-step at the higher level, and the process unfolds again. One can concisely summarise this by saying that a state at the higher level corresponds to a trial of observation epochs at the lower level. Of course, there is no limit to the number of layers one can stack in a hierarchical model.

To obtain the associated belief updating, one computes free energy at the lower level by conditioning the probability distributions from Bayes rule by the variables from the higher levels. This means that one performs belief updating at the lower levels independently of the higher levels. Then, one computes the variational free energy at the higher levels by treating the lower levels as outcomes. For more details on the specificities of the scheme see [3].

B Expected free energy

At the heart of active inference is a description of a certain class of systems at non-equilibrium steady-state (NESS) [68, 69]. An important consequence of NESS is the existence of a steady-state probability distribution $P(s_\tau, A)$ that the agent is guaranteed to reach given a sufficient amount of time. Intuitively, this distribution should be thought as the agent's preferences over states and model parameters. Practically, this means that the agent selects policies, such that its predicted states $Q(s_\tau, A)$ at some future time point $\tau > t$ – usually, the time horizon of a policy T – reach its preferences $P(s_\tau, A)$, which are specified by the generative model. In the following, we will show how a specific family of distributions $Q(\pi)$ guarantee an agent to reach its preferences. Then, we will see how NESS enables in fact to extract one single canonical member of this family: the (softmax negative) *expected free energy*.

Objective: we seek distributions over policies that imply steady-state solutions; i.e., when the final distribution does not depend upon initial observations. Such solutions ensure that, on average, stochastic policies lead to a steady-state or target distribution specified by the generative model. These solutions exist in virtue of conditional independencies, where the hidden states provide a Markov blanket that separates policies from outcomes. In other words, policies cause final states that cause outcomes. In what follows, $\tau > t$ is a future time and $Q := Q(o_\tau, s_\tau, A, \pi) \approx P(o_\tau, s_\tau, A, \pi | o_{1:t})$ is the corresponding approximate posterior distribution, given initial conditions $o_{1:t}$.

Lemma 1 (Steady-state). *The surprisal over policies $-\log Q(\pi)$ and the Gibbs energy,*

$$G(\pi; \beta) = D_{KL}[Q(s_\tau, A|\pi)||P(s_\tau, A)] - \mathbb{E}_{Q(o_\tau, s_\tau, A|\pi)}[\beta \log P(o_\tau|s_\tau, A)] \tag{34}$$

$$\beta := \frac{\mathbb{E}_Q[\log Q(\pi|s_\tau, A)]}{\mathbb{E}_Q[\log P(o_\tau|s_\tau, A)]} \geq 0 \tag{35}$$

are equal on average under Q if and only if the system reaches steady-state. Explicitly:

$$\mathbb{E}_Q[-\log Q(\pi)] = \mathbb{E}_Q[G(\pi; \beta)] \iff D_{KL}[Q(s_\tau, A)||P(s_\tau, A)] = 0 \tag{36}$$

Here, $\beta \geq 0$ characterises the steady-state with the relative precision (i.e., negative entropy) of policies and final outcomes, given final states. The generative model stipulates steady-state, in the sense that distribution over final states (and outcomes) does not depend upon initial observations. Here, the generative and predictive distributions simply express the conditional independence between policies and final outcomes, given final states. Note that when $\beta = 1$, Gibbs energy becomes expected free energy.

Proof. Let us unpack the Gibbs energy expected under Q:

$$\begin{aligned}
\mathbb{E}_Q[G(\pi; \beta)] &= \mathbb{E}_Q[D_{KL}[Q(s_\tau, A|\pi)||P(s_\tau, A)] - \mathbb{E}_{Q(o_\tau, s_\tau, A|\pi)}[\beta \log P(o_\tau|s_\tau, A)]] \\
&= \mathbb{E}_Q[D_{KL}[Q(s_\tau, A|\pi)||P(s_\tau, A)]] \\
&\quad - \frac{\mathbb{E}_Q[\log Q(\pi|s_\tau, A)]}{\mathbb{E}_Q[\log P(o_\tau|s_\tau, A)]} \mathbb{E}_Q[\mathbb{E}_{Q(o_\tau, s_\tau, A|\pi)}[\log P(o_\tau|s_\tau, A)]] \\
&= \mathbb{E}_Q[\log Q(s_\tau, A|\pi) - \log P(s_\tau, A) - \log Q(\pi|s_\tau, A)] \\
&= \mathbb{E}_Q[-\log Q(\pi) - \log P(s_\tau, A) + \log Q(s_\tau, A)] \\
&= \mathbb{E}_Q[-\log Q(\pi)] + D_{KL}[Q(s_\tau, A)||P(s_\tau, A)]
\end{aligned} \tag{37}$$

And the result is immediate. □

A straightforward consequence of Lemma 1, is that each distribution

$$Q(\pi) = \sigma(-G(\pi; \beta)), \quad \beta \geq 0 \tag{38}$$

describes a certain kind of system that self-organises to some steady-state distribution. This family of distributions has interesting interpretations: for example, the case $\beta = 0$ corresponds to standard stochastic control, variously known as KL control or risk-sensitive control [142]:

$$G(\pi; 0) = D_{KL}[Q(s_\tau, A|\pi)||P(s_\tau, A)] \geq D_{KL}[Q(s_\tau|\pi)||P(s_\tau)] \tag{39}$$

In other words, one chooses policies that minimise the KL divergence between the predictive and target distribution. More generally, when $\beta > 0$, policies are more likely when they simultaneously minimise the entropy of outcomes, given states. In other words, $\beta > 0$ ensures that the system exhibits itinerant behaviour. One can see that KL control may arise in this case if the entropy of the likelihood mapping remains constant with respect to policies.

Remark 2. *It is possible to extend this framework by considering systems that reach their preferences at a collection of time-steps into the future, say $\tau_1, ..., \tau_n > t$. In this case, one can adapt the proof of Lemma 1 to obtain:*

$$\mathbb{E}_Q \left[\sum_{i=1}^{n} G(\pi, \tau_i; \beta) \right] = \mathbb{E}_Q[-n \log Q(\pi)] + \sum_{i=1}^{n} D_{KL}[Q(s_{\tau_i}, A)||P(s_{\tau_i}, A)] \tag{40}$$

where $G(\pi, \tau_i; \beta)$ is the Gibbs free energy of Lemma 1, replacing τ by τ_i. In this case, the canonical choice of approximate posterior over policies would be:

$$Q(\pi) = \sigma \left(\frac{1}{n} \sum_{i=1}^{n} G(\pi, \tau_i; \beta) \right) \tag{41}$$

One perspective – on the distinction between simple and general steady-states – is in terms of uncertainty about policies. For example, simple steady-states preclude uncertainty about which policy led to a final state. This would be appropriate for describing classical systems (that follow a unique path of least action), where it would be possible to infer which policy had been pursued, given the initial and final outcomes. Conversely, in general steady-state systems (e.g., mice, Homo sapiens), simply knowing that 'you are here' does not tell me 'how you got here', even if I knew where you were this morning. Put another way, there are lots of paths or policies open to systems that attain a general steady state.

In active inference, we are interested in a certain class of systems that self-organise to general steady-states; namely, those that move through a large number of probabilistic configurations from their initial state to their final steady-state. The treatments in [68, 69] effectively turn the steady-state lemma on its head by assuming NESS is stipulatively true – and then characterise the ensuing self-organisation in terms of Bayes optimal policies:

Corollary 3 (Active inference [69]). *If a system attains a general steady-state, it will appear to behave in a Bayes optimal fashion – both in terms of optimal Bayesian design (i.e., exploration) and Bayesian decision theory (i.e., exploitation). Crucially, the loss function defining Bayesian risk is the negative log evidence for the generative model entailed by an agent. In short, systems (i.e., agents) that attain general steady-states will look as if they are responding to epistemic affordances [44].*

So far, we have deduced the distribution over policies of systems that reach steady-state. However, recall that reaching steady-state is only a consequence of NESS. In fact, NESS dynamics under a Markov blanket (c.f., Figure 1) imply a slightly stronger statement: the most likely trajectories of systems described by active inference are those which minimise *expected free energy* [68, 69] – this is exactly the case $\beta = 1$ in (38). This is nice, since many existing theories of cognition and control emerge under this specific imperative (c.f., Figure 4).

C Computing expected free energy

In this appendix, we present the derivations underlying the analytical expression of the expected free energy that is used in `spm_MDP_VB_X.m`. Following [119], we can reexpress the expected free energy in the following form:

$$G(\pi) \quad = \quad \underbrace{\mathbb{E}_{Q(s_\tau|\pi)}[\mathrm{H}[P(o_\tau|s_\tau)]]}_{\text{Ambiguity}} + \underbrace{D_{KL}[Q(s_\tau|\pi)||P(s_\tau)]}_{\text{Risk (states)}} - \underbrace{\mathbb{E}_{P(o_\tau|s_\tau)Q(s_\tau|\pi)}[D_{KL}[Q(A|o_\tau,s_\tau)||Q(A)]]}_{\text{Novelty}} \quad (42)$$

Here, $Q(A|o_\tau, s_\tau)$ denotes approximate posterior beliefs about A if we knew occurence of the state outcome pair (o_τ, s_τ). In the following, we show that we can compute the expected free energy in the following way

$$G(\pi) \approx \underbrace{H \cdot \mathbf{s}_{\pi\tau}}_{\text{Ambiguity}} + \underbrace{\mathbf{s}_{\pi\tau} \cdot (\log \mathbf{s}_{\pi\tau} - \log C)}_{\text{Risk (states)}} - \underbrace{\mathbf{As}_{\pi\tau} \cdot W\mathbf{s}_{\pi\tau}}_{\text{Novelty}}$$

$$H := -\mathrm{diag}[\mathbf{A} \cdot \log \mathbf{A}] \tag{43}$$

$$W := \frac{1}{2}\left(\mathbf{a}^{\odot(-1)} - \mathbf{a}_0^{\odot(-1)}\right)$$

when the agent's preferences C are expressed in terms of preferences over states. When preferences are expressed in terms of outcomes (as is currently implemented in `spm_MDP_VB_X.m`), the risk term instead becomes

$$\underbrace{(\mathbf{As}_{\pi\tau}) \cdot (\log(\mathbf{As}_{\pi\tau}) - \log C)}_{\text{Risk (outcomes)}} \tag{44}$$

C.1 Ambiguity

The ambiguity term of (42) is $\mathbb{E}_{Q(s_\tau|\pi)}[\mathrm{H}[P(o_\tau|s_\tau)]]$. By definition, the entropy inside the expectation is:

$$\mathrm{H}[P(o_\tau|s_\tau)] = -\sum_{o_\tau \in O} P(o_\tau|s_\tau) \log P(o_\tau|s_\tau) \tag{45}$$

The first factor inside the sum corresponds to:

$$
\begin{aligned}
P(o_\tau|s_\tau) &= \int P(o_\tau, A|s_\tau)\, dA \\
&= \int P(o_\tau|s_\tau, A)P(A)\, dA \\
&= \int o_\tau \cdot As_\tau P(A)\, dA \\
&\approx \int o_\tau \cdot As_\tau Q(A)\, dA \\
&= o_\tau \cdot \mathbb{E}_{Q(A)}[A]s_\tau \\
&= o_\tau \cdot \mathbf{A}s_\tau
\end{aligned}
\tag{46}
$$

Here we have replaced the prior over model parameters $P(A)$ by the approximate posterior $Q(A)$. This is not necessary, but in numerical simulations since learning occurs once at the end of the trial the two can be interchanged – furthermore, this allows us to reuse previously introduced notation. In any case, this tells us that the entropy can be re-expressed as:

$$
\begin{aligned}
\mathrm{H}[P(o_\tau|s_\tau)] &= -\sum_{o_\tau \in O} (o_\tau \cdot \mathbf{A}s_\tau)(o_\tau \cdot \log(\mathbf{A})s_\tau) \\
&= -\sum_{i=1}^{n} (\mathbf{A}_{\bullet i}s_\tau)(\log(\mathbf{A}_{\bullet i})s_\tau) \\
&= -\sum_{i=1}^{n} (\mathbf{A}_{\bullet i} \odot \log(\mathbf{A}_{\bullet i}))s_\tau \\
&= -(\mathbf{A} \odot \log \mathbf{A})s_\tau \\
&= -\mathrm{diag}[\mathbf{A} \cdot \log \mathbf{A}] \cdot s_\tau
\end{aligned}
\tag{47}
$$

27

Finally,

$$\underbrace{\mathbb{E}_{Q(s_\tau|\pi)}[\mathbf{H}[P(o_\tau|s_\tau)]]}_{\text{Ambiguity}} = H \cdot \mathbf{s}_{\pi\tau} \tag{48}$$

$$H := -\text{diag}[\mathbf{A} \cdot \log \mathbf{A}]$$

C.2 Risk

The risk term of (42) is the KL divergence between predicted states following a particular policy and preferred states. This can be expressed as:

$$\underbrace{D_{KL}[Q(s_\tau|\pi)||P(s_\tau)]}_{\text{Risk (states)}} = \mathbf{s}_{\pi\tau} \cdot (\mathbf{s}_{\pi\tau} - \log C) \tag{49}$$

Where the vector $C \in \mathbb{R}^m$ encodes preference over states $P(s_\tau) = Cat(C)$. However, it is also possible to approximate this risk term over states by a risk term over outcomes (c.f., (15)), as is currently implemented in `spm_MDP_VB_X.m`. In this case, if $C \in \mathbb{R}^n$ denotes the preferences over outcomes $P(o_\tau) = Cat(C)$:

$$\underbrace{D_{KL}[Q(o_\tau|\pi)||P(o_\tau)]}_{\text{Risk (outcomes)}} = (\mathbf{A}\mathbf{s}_{\pi\tau}) \cdot (\log(\mathbf{A}\mathbf{s}_{\pi\tau}) - \log C) \tag{50}$$

C.3 Novelty

The novelty term of (42) is $\mathbb{E}_{P(o_\tau|s_\tau)Q(s_\tau|\pi)}[D_{KL}[Q(A|o_\tau,s_\tau)||Q(A)]]$ where

$$Q(A) = \prod_{i=1}^{m} Q(A_{\bullet i}), \quad Q(A_{\bullet i}) = Dir(\mathbf{a}_{\bullet i}) \tag{51}$$

$$Q(A|o_\tau,s_\tau) = \prod_{i=1}^{m} Q(A_{\bullet i}|o_\tau,s_\tau), \quad Q(A_{\bullet i}|o_\tau,s_\tau) := Dir(\mathbf{a}'_{\bullet i}) \tag{52}$$

The KL divergence between both distributions (c.f., (18)) can be expressed as:

$$\begin{aligned}
D_{KL}[Q(A|o_\tau,s_\tau)||Q(A)] = \sum_{i=1}^{m}[\log\Gamma(\mathbf{a}'_{0i}) - \sum_{k=1}^{n}\log\Gamma(\mathbf{a}'_{ki}) - \log\Gamma(\mathbf{a}_{0i}) \\
+ \sum_{k=1}^{n}\log\Gamma(\mathbf{a}_{ki})] + (\mathbf{a}' - \mathbf{a}) \cdot (\psi(\mathbf{a}') - \psi(\mathbf{a}'_0))
\end{aligned} \tag{53}$$

where ψ is the digamma function. We now want to make sense of $\mathbf{a}'$. Suppose that at time τ the agents knows the possible outcome j and possible state k as in $Q(A|o_\tau,s_\tau)$ (c.f., Table 2 for terminology). This means that in this case, beliefs about hidden states correspond to the true state; in other words, $\mathbf{s}_\tau = s_\tau$. We can then use the rule of accumulation of Dirichlet parameters to deduce $\mathbf{a}' = \mathbf{a} + o_\tau \otimes s_\tau$. In other words, $\mathbf{a}'_{jk} = \mathbf{a}_{jk} + 1$ and the remaining components are identical. Using the well-known identity:

$$\Gamma(x+1) = x\Gamma(x) \Rightarrow \log\Gamma(x+1) = \log x + \log\Gamma(x) \tag{54}$$

we can compute (53):

$$\begin{aligned}
D_{KL}[Q(A|o_\tau,s_\tau)||Q(A)] = \log\Gamma(\mathbf{a}_{0k}+1) - \log\Gamma(\mathbf{a}_{0k}) - \log\Gamma(\mathbf{a}_{jk}+1) + \log\Gamma(\mathbf{a}_{jk}) + \psi(\mathbf{a}_{jk}+1) - \psi(\mathbf{a}_{0k} \\
= \log\mathbf{a}_{0k} - \log\mathbf{a}_{jk} + \psi(\mathbf{a}_{jk}+1) - \psi(\mathbf{a}_{0k}+1)
\end{aligned} \tag{55}$$

Using the definition of the digamma function $\psi(x) = \frac{d}{dx}\log\Gamma(x)$ we obtain:

$$
\begin{aligned}
D_{KL}[Q(A|o_\tau, s_\tau)||Q(A)] &= \log\mathbf{a}_{0k} - \log\mathbf{a}_{jk} + \frac{d}{d\mathbf{a}_{jk}}(\log\Gamma(\mathbf{a}_{jk}+1)) - \frac{d}{d\mathbf{a}_{0k}}(\log\Gamma(\mathbf{a}_{0k}+1)) \\
&= \log\mathbf{a}_{0k} - \log\mathbf{a}_{jk} + \frac{d}{d\mathbf{a}_{jk}}(\log\Gamma(\mathbf{a}_{jk}+1)) - \frac{d}{d\mathbf{a}_{0k}}(\log\mathbf{a}_{0k} + \log\Gamma(\mathbf{a}_{0k})) \quad (56) \\
&= \log\mathbf{a}_{0k} - \log\mathbf{a}_{jk} + \frac{1}{\mathbf{a}_{jk}} - \frac{1}{\mathbf{a}_{0k}} + \psi(\mathbf{a}_{jk}) - \psi(\mathbf{a}_{0k})
\end{aligned}
$$

We can use an asymptotic expansion of the digamma function to simplify the expression:

$$
\begin{aligned}
\psi(x) &\approx \log x - \frac{1}{2x} + \cdots \\
\Rightarrow D_{KL}[Q(A|o_\tau, s_\tau)||Q(A)] &\approx \frac{1}{2\mathbf{a}_{jk}} - \frac{1}{2\mathbf{a}_{0k}}
\end{aligned}
\quad (57)
$$

Finally, the analytical expression of the novelty term:

$$
\begin{aligned}
\mathbb{E}_{P(o_\tau|s_\tau)Q(s_\tau|\pi)}[D_{KL}[Q(A|o_\tau, s_\tau)||Q(A)]] &\approx \mathbf{A}\mathbf{s}_{\pi\tau} \cdot W\mathbf{s}_{\pi\tau} \\
W &:= \frac{1}{2}\left(\mathbf{a}^{\odot-1} - \mathbf{a}_0^{\odot-1}\right)
\end{aligned}
\quad (58)
$$

References

[1] Raphael Kaplan and Karl J. Friston. Planning and navigation as active inference. *Biological Cybernetics*, 112(4):323–343, August 2018.

[2] Karl J. Friston, Richard Rosch, Thomas Parr, Cathy Price, and Howard Bowman. Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews*, 90:486–501, July 2018.

[3] Karl J. Friston, Marco Lin, Christopher D. Frith, Giovanni Pezzulo, J. Allan Hobson, and Sasha Ondobaka. Active Inference, Curiosity and Insight. *Neural Computation*, 29(10):2633–2683, October 2017.

[4] Thomas Parr and Karl J. Friston. Active inference and the anatomy of oculomotion. *Neuropsychologia*, 111:334–343, March 2018.

[5] M. Berk Mirza, Rick A. Adams, Christoph D. Mathys, and Karl J. Friston. Scene Construction, Visual Foraging, and Active Inference. *Frontiers in Computational Neuroscience*, 10, June 2016.

[6] Thomas Parr and Karl J. Friston. Uncertainty, epistemics and active inference. *Journal of the Royal Society Interface*, 14(136), November 2017.

[7] Thomas Parr and Karl J. Friston. The Computational Anatomy of Visual Neglect. *Cerebral Cortex (New York, N.Y.: 1991)*, 28(2):777–790, January 2018.

[8] Rick A. Adams, Klaas Enno Stephan, Harriet R. Brown, Christopher D. Frith, and Karl J. Friston. The Computational Anatomy of Psychosis. *Frontiers in Psychiatry*, 4, 2013.

[9] Jelle Bruineberg, Erik Rietveld, Thomas Parr, Leendert van Maanen, and Karl J Friston. Free-energy minimization in joint agent-environment systems: A niche construction perspective. *Journal of Theoretical Biology*, 455:161–178, October 2018.

[10] A Constant, M Ramstead, Samuel P. L. Veissière, J O Campbell, and K.J. Friston. A variational approach to niche construction. *Journal of The Royal Society Interface*, 15(141):20170685, April 2018.

[11] Axel Constant, Maxwell J. D. Ramstead, Samuel P. L. Veissière, and Karl Friston. Regimes of Expectations: An Active Inference Model of Social Conformity and Human Decision Making. *Frontiers in Psychology*, 10, 2019.

[12] M. Berk Mirza, Rick A. Adams, Thomas Parr, and Karl Friston. Impulsivity and Active Inference. *Journal of Cognitive Neuroscience*, 31(2):202–220, February 2019.

[13] Beren Millidge. Implementing Predictive Processing and Active Inference: Preliminary Steps and Results. Preprint, PsyArXiv, March 2019.

[14] Ozan Çatal, Johannes Nauta, Tim Verbelen, Pieter Simoens, and Bart Dhoedt. Bayesian policy selection using active inference. *arXiv:1904.08149 [cs]*, April 2019.

[15] Karl J. Friston, Jean Daunizeau, and Stefan J. Kiebel. Reinforcement Learning or Active Inference? *PLoS ONE*, 4(7):e6421, July 2009.

[16] Karl Friston, Rick Adams, and Read Montague. What is value—accumulated reward or evidence? *Frontiers in Neurorobotics*, 6, 2012.

[17] Rafal Bogacz. A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, 76:198–211, February 2017.

[18] Christopher L. Buckley, Chang Sub Kim, Simon McGregor, and Anil K. Seth. The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, 81:55–79, December 2017.

[19] S. J. Luck and E. K. Vogel. The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657):279–281, November 1997.

[20] Weiwei Zhang and Steven J. Luck. Discrete Fixed-Resolution Representations in Visual Working Memory. *Nature*, 453(7192):233–235, May 2008.

[21] Howard Eichenbaum, Paul Dudchenko, Emma Wood, Matthew Shapiro, and Heikki Tanila. The Hippocampus, Memory, and Place Cells: Is It Spatial Memory or a Memory Space? *Neuron*, 23(2):209–226, June 1999.

[22] J. O'Keefe and J. Dostrovsky. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1):171–175, November 1971.

[23] Nathaniel D. Daw, John P. O'Doherty, Peter Dayan, Ben Seymour, and Raymond J. Dolan. Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095):876–879, June 2006.

[24] Paul Reverdy, Vaibhav Srivastava, and Naomi E. Leonard. Modeling Human Decision-making in Generalized Gaussian Multi-armed Bandits. *arXiv:1307.6134 [cs, math, stat]*, July 2013.

[25] Charley M. Wu, Eric Schulz, Maarten Speekenbrink, Jonathan D. Nelson, and Björn Meder. Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour*, 2(12):915–924, December 2018.

[26] Nathaniel D. Daw, Samuel J. Gershman, Ben Seymour, Peter Dayan, and Raymond J. Dolan. Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*, 69(6):1204–1215, March 2011.

[27] Karl J. Friston, Thomas Parr, and Bert de Vries. The graphical brain: Belief propagation and active inference. *Network Neuroscience*, 1(4):381–414, December 2017.

[28] Thomas Parr and Karl J. Friston. The Discrete and Continuous Brain: From Decisions to Movement—And Back Again. *Neural Computation*, 30(9):2319–2347, September 2018.

[29] Thomas Parr and Karl J. Friston. The computational pharmacology of oculomotion. *Psychopharmacology*, April 2019.

[30] Karl Friston, Spyridon Samothrakis, and Read Montague. Active inference and agency: Optimal control without cost functions. *Biological Cybernetics*, 106(8):523–541, October 2012.

[31] Thomas H. B. FitzGerald, Philipp Schwartenbeck, Michael Moutoussis, Raymond J. Dolan, and Karl Friston. Active inference, evidence accumulation, and the urn task. *Neural Computation*, 27(2):306–328, February 2015.

[32] Thomas H. B. FitzGerald, Rosalyn J. Moran, Karl J. Friston, and Raymond J. Dolan. Precision and neuronal dynamics in the human posterior parietal cortex during evidence accumulation. *NeuroImage*, 107:219–228, February 2015.

[33] Philipp Schwartenbeck, Thomas H. B. FitzGerald, Christoph Mathys, Ray Dolan, Friedrich Wurst, Martin Kronbichler, and Karl Friston. Optimal inference with suboptimal models: Addiction and active Bayesian inference. *Medical Hypotheses*, 84(2):109–117, February 2015.

[34] Thomas H. B. FitzGerald, Raymond J. Dolan, and Karl Friston. Dopamine, reward learning, and active inference. *Frontiers in Computational Neuroscience*, 9, November 2015.

[35] Karl Friston, Philipp Schwartenbeck, Thomas FitzGerald, Michael Moutoussis, Timothy Behrens, and Raymond J. Dolan. The anatomy of choice: Dopamine and decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655), November 2014.

[36] Philipp Schwartenbeck, Thomas H. B. FitzGerald, Christoph Mathys, Ray Dolan, and Karl Friston. The Dopaminergic Midbrain Encodes the Expected Certainty about Desired Outcomes. *Cerebral Cortex (New York, N.Y.: 1991)*, 25(10):3434–3445, October 2015.

[37] Philipp Schwartenbeck, Thomas H. B. FitzGerald, Christoph Mathys, Ray Dolan, Martin Kronbichler, and Karl Friston. Evidence for surprise minimization over value maximization in choice behavior. *Scientific Reports*, 5:16575, November 2015.

[38] Michael Moutoussis, Nelson J. Trujillo-Barreto, Wael El-Deredy, Raymond J. Dolan, and Karl J. Friston. A formal model of interpersonal inference. *Frontiers in Human Neuroscience*, 8, March 2014.

[39] Karl Friston, Francesco Rigoli, Dimitri Ognibene, Christoph Mathys, Thomas Fitzgerald, and Giovanni Pezzulo. Active inference and epistemic value. *Cognitive Neuroscience*, 6(4):187–214, October 2015.

[40] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, John O'Doherty, and Giovanni Pezzulo. Active inference and learning. *Neuroscience & Biobehavioral Reviews*, 68:862–879, September 2016.

[41] Thomas H. B. FitzGerald, Raymond J. Dolan, and Karl J. Friston. Model averaging, optimal inference, and habit formation. *Frontiers in Human Neuroscience*, 8, 2014.

[42] Karl Friston and Gyorgy Buzsáki. The Functional Anatomy of Time: What and When in the Brain. *Trends in Cognitive Sciences*, 20(7):500–511, July 2016.

[43] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, and Giovanni Pezzulo. Active Inference: A Process Theory. *Neural Computation*, 29(1):1–49, January 2017.

[44] Thomas Parr and Karl J Friston. Working memory, attention, and salience in active inference. *Scientific Reports*, 7(1):14678, December 2017.

[45] Thomas Parr, Geraint Rees, and Karl J. Friston. Computational Neuropsychology and Bayesian Inference. *Frontiers in Human Neuroscience*, 12, 2018.

[46] Thomas Parr, Rajeev Vijay Rikhye, Michael M Halassa, and Karl J Friston. Prefrontal Computation as Active Inference. *Cerebral Cortex*, page bhz118 e0190429 e12112 e37454, July 2019.

[47] David Benrimoh, Thomas Parr, Peter Vincent, Rick A. Adams, and Karl Friston. Active Inference and Auditory Hallucinations. *Computational Psychiatry*, 2:183–204, November 2018.

[48] Thomas Parr, David A. Benrimoh, Peter Vincent, and Karl J. Friston. Precision and False Perceptual Inference. *Frontiers in Integrative Neuroscience*, 12, 2018.

[49] Anna C. Sales, Karl J. Friston, Matthew W. Jones, Anthony E. Pickering, and Rosalyn J. Moran. Locus Coeruleus tracking of prediction errors optimises cognitive flexibility: An Active Inference model. *bioRxiv*, page 340620, June 2018.

[50] Peter Vincent, Thomas Parr, David Benrimoh, and Karl J. Friston. With an eye on uncertainty: Modelling pupillary responses to environmental volatility. *PLOS Computational Biology*, 15(7):e1007126, 05-Jul-2019.

[51] Maell Cullen, Ben Davey, Karl J. Friston, and Rosalyn J. Moran. Active Inference in OpenAI Gym: A Paradigm for Computational Investigations Into Psychiatric Illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(9):809–818, September 2018.

[52] Alexander Tschantz, Manuel Baltieri, Anil K. Seth, and Christopher L. Buckley. Scaling active inference. *arXiv:1911.10601 [cs, eess, math, stat]*, November 2019.

[53] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, 2006.

[54] Yang Xitong. Understanding the Variational Lower Bound. 2017.

[55] Biswa Sengupta and Karl Friston. Approximate Bayesian inference as a gauge theory. In *Computational Biology Workshop*, volume 14, page e1002400, March 2016.

[56] Biswa Sengupta, Arturo Tozzi, Gerald K. Cooray, Pamela K. Douglas, and Karl J. Friston. Towards a Neuronal Gauge Theory. *PLOS Biology*, 14(3):e1002400, 08-Mar-2016.

[57] Laurence Aitchison and Máté Lengyel. With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, 46:219–227, October 2017.

[58] Karl Friston. The history of the future of the Bayesian brain. *NeuroImage*, 62(2):1230–1233, August 2012.

[59] David C. Knill and Alexandre Pouget. The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712–719, December 2004.

[60] Ryan T. McKay and Daniel C. Dennett. The evolution of misbelief. *The Behavioral and Brain Sciences*, 32(6):493–510; discussion 510–561, December 2009.

[61] Tali Sharot. The optimism bias. *Current Biology*, 21(23):R941–R945, December 2011.

[62] Karl Friston. The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, 13(7):293–301, July 2009.

[63] Karl Friston. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, February 2010.

[64] Karl J. Friston and Klaas E. Stephan. Free-energy and the brain. *Synthese*, 159(3):417–458, November 2007.

[65] Karl Friston, James Kilner, and Lee Harrison. A free energy principle for the brain. *Journal of Physiology-Paris*, 100(1-3):70–87, July 2006.

[66] J. M. Fuster. Prefrontal cortex and the bridging of temporal gaps in the perception-action cycle. *Annals of the New York Academy of Sciences*, 608:318–329; discussion 330–336, 1990.

[67] Jakob Hohwy. The Self-Evidencing Brain. *Noûs*, 50(2):259–285, June 2016.

[68] Thomas Parr, Lancelot Da Costa, and Karl J. Friston. Markov blankets, information geometry and stochastic thermodynamics. *Phil. Trans. R. Soc. A.*, 2019.

[69] Karl Friston. A free energy principle for a particular physics. *BioArxiv*, page 148, 2019.

[70] Michael Kirchhoff, Thomas Parr, Ensor Palacios, Karl Friston, and Julian Kiverstein. The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of The Royal Society Interface*, 15(138):20170792, January 2018.

[71] Judea Pearl. Graphical Models for Probabilistic and Causal Reasoning. In Philippe Smets, editor, *Quantified Representation of Uncertainty and Imprecision*, Handbook of Defeasible Reasoning and Uncertainty Management Systems, pages 367–389. Springer Netherlands, Dordrecht, 1998.

[72] Matthew James Beal. Variational Algorithms for Approximate Bayesian Inference. page 281, 2003.

[73] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518):859–877, April 2017.

[74] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. An Introduction to Variational Methods for Graphical Models. In Michael I. Jordan, editor, *Learning in Graphical Models*, pages 105–161. Springer Netherlands, Dordrecht, 1998.

[75] K J Aström. Optimal Control of Markov Processes with Incomplete State Information. *Journal of Mathematical Analysis and Applications*, 10, 1965.

[76] Andrew Barto and Richard Sutton. *Reinforcement Learning: An Introduction*. 1992.

[77] James V Stone. *Artificial Intelligence Engines: A Tutorial Introduction to the Mathematics of Deep Learning*. 2019.

[78] C. C. White. Markov decision processes. In *Encyclopedia of Operations Research and Management Science*, pages 484–486. Springer US, Boston, MA, 2001.

[79] G E P Box and George C Tiao. Multiparameter problems from a Bayesian point of view. *The Annals of Mathematical Statistics*, 1965.

[80] Greg M Allenby, Peter E Rossi, and Robert E McCulloch. Hierarchical Bayes Models: A Practitioners Guide. 2005.

[81] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. 1988.

[82] Hans-Andrea Loeliger. An introduction to factor graphs. *IEEE Signal Processing Magazine*, 2004.

[83] Toshiyuki Tanaka. A Theory of Mean Field Approximation. page 10, 1999.

[84] Thomas Parr, Dimitrije Markovic, Stefan J. Kiebel, and Karl J. Friston. Neuronal message passing using Mean-field, Bethe, and Marginal approximations. *Scientific Reports*, 9(1):1889, December 2019.

[85] Sarah Schwöbel, Stefan Kiebel, and Dimitrije Marković. Active Inference, Belief Propagation, and the Bethe Approximation. *Neural Computation*, 30(9):2530–2567, September 2018.

[86] James Tee and Desmond P. Taylor. Is Information in the Brain Represented in Continuous or Discrete Form? *arXiv:1805.01631 [cs, math, q-bio]*, May 2018.

[87] Timothy J. Buschman and Earl K. Miller. Shifting the Spotlight of Attention: Evidence for Discrete Computations in Cognition. *Frontiers in Human Neuroscience*, 4, 2010.

[88] J. Duncan, R. Ward, and K. Shapiro. Direct measurement of attentional dwell time in human vision. *Nature*, 369(6478):313–315, May 1994.

[89] Ayelet Nina Landau and Pascal Fries. Attention Samples Stimuli Rhythmically. *Current Biology*, 22(11):1000–1004, June 2012.

[90] Simon Hanslmayr, Gregor Volberg, Maria Wimber, Sarang S. Dalal, and Mark W. Greenlee. Prestimulus Oscillatory Phase at 7 Hz Gates Cortical Information Flow and Visual Perception. *Current Biology*, 23(22):2273–2278, November 2013.

[91] E T Rolls and M J Tovee. Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 257(1348):9–15, July 1994.

[92] Karl Friston. Hierarchical Models in the Brain. *PLoS Computational Biology*, 4(11):e1000211, November 2008.

[93] Zenas C. Chao, Kana Takaura, Liping Wang, Naotaka Fujii, and Stanislas Dehaene. Large-Scale Cortical Networks for Hierarchical Prediction and Prediction Error in the Primate Brain. *Neuron*, 100(5):1252–1266.e3, May 2018.

[94] Sandra Iglesias, Christoph Mathys, Kay H. Brodersen, Lars Kasper, Marco Piccirelli, Hanneke E. M. den Ouden, and Klaas E. Stephan. Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. *Neuron*, 80(2):519–530, October 2013.

[95] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951.

[96] Horace Barlow. Redundancy reduction revisited. *Comput. Neural Syst.*, page 13, 2001.

[97] Yang Dan, Joseph J. Atick, and R. Clay Reid. Efficient Coding of Natural Scenes in the Lateral Geniculate Nucleus: Experimental Test of a Computational Theory. *Journal of Neuroscience*, 16(10):3351–3362, May 1996.

[98] Michael S. Lewicki. Efficient coding of natural sounds. *Nature Neuroscience*, 5(4):356–363, April 2002.

[99] Bruno A Olshausen and David J Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4):481–487, August 2004.

[100] Bruno A. Olshausen and Kevin N. O'Connor. A new window on sound. *Nature Neuroscience*, 5(4):292–294, April 2002.

[101] Emanuel Todorov and Michael I. Jordan. Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11):1226–1235, November 2002.

[102] Peter Bossaerts and Carsten Murawski. From behavioural economics to neuroeconomics to decision neuroscience: The ascent of biology in research on human decision making. *Current Opinion in Behavioral Sciences*, 5:37–42, October 2015.

[103] L. M. Optican and B. J. Richmond. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *Journal of Neurophysiology*, 57(1):162–178, January 1987.

[104] R Linsker. Perceptual Neural Organization: Some Approaches Based on Network Models and Information Theory. *Annual Review of Neuroscience*, 13(1):257–281, 1990.

[105] H. B. Barlow. *Possible Principles Underlying the Transformations of Sensory Messages*. The MIT Press, 1961.

[106] Hermann Haken. *Synergetics: An Introduction Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology*. Springer Series in Synergetics. Springer-Verlag, Berlin Heidelberg, 2 edition, 1978.

[107] Stuart A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993.

[108] G. Nicolis and I. Prigogine. *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations*. Wiley-Blackwell, New York, June 1977.

[109] W. R. Ashby. Principles of the Self-Organizing Dynamic System. *The Journal of General Psychology*, 37(2):125–128, October 1947.

[110] Claude Bernard. *Lectures on the Phenomena of Life Common to Animals and Plants*. Thomas, 1974.

[111] Roger C Conant and W. R. Ashby. Every good regulator of a system must be a model of that system. *Int. J. Systems Sci.*, 1(2):89–97, 1970.

[112] David J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge, UK ; New York, sixth printing 2007 edition edition, September 2003.

[113] David MacKay. A Free Energy Minimization Algorithm for Decoding and Cryptanalysis. In *Electronic Letters*, 1995.

[114] Hermann von Helmholtz and James P. C Southall. *Helmholtz's Treatise on Physiological Optics*. Dover Publications, New York, 1962. OCLC: 523553.

[115] R. L. Gregory. Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 290(1038):181–197, July 1980.

[116] Peter Dayan, Geoffrey E. Hinton, Radford M. Neal, and Richard S. Zemel. The Helmholtz Machine. *Neural Computation*, 7(5):889–904, September 1995.

[117] J.S. Yedidia, W.T. Freeman, and Y. Weiss. Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, July 2005.

[118] T. Heskes. Convexity Arguments for Efficient Minimization of the Bethe and Kikuchi Free Energies. *Journal of Artificial Intelligence Research*, 26:153–190, June 2006.

[119] Thomas Parr. *The Computational Neurology of Active Vision*. Ph.D., University College London, London, 2019.

[120] Kaare Brandt Petersen and Michael Syskind Pedersen. The Matrix Cookbook. page 72.

[121] Mikael Lundqvist, Jonas Rose, Pawel Herman, Scott L. Brincat, Timothy J. Buschman, and Earl K. Miller. Gamma and Beta Bursts Underlie Working Memory. *Neuron*, 90(1):152–164, April 2016.

[122] Rajesh P. N. Rao and Dana H. Ballard. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, January 1999.

[123] Andre M. Bastos, W. Martin Usrey, Rick A. Adams, George R. Mangun, Pascal Fries, and Karl J. Friston. Canonical Microcircuits for Predictive Coding. *Neuron*, 76(4):695–711, November 2012.

[124] Karl Friston, Jérémie Mattout, Nelson Trujillo-Barreto, John Ashburner, and Will Penny. Variational free energy and the Laplace approximation. *NeuroImage*, 34(1):220–234, January 2007.

[125] André C. Marreiros, Jean Daunizeau, Stefan J. Kiebel, and Karl J. Friston. Population dynamics: Variance and the sigmoid activation function. *NeuroImage*, 42(1):147–157, August 2008.

[126] Gustavo Deco, Viktor K. Jirsa, Peter A. Robinson, Michael Breakspear, and Karl Friston. The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields. *PLoS Computational Biology*, 4(8):e1000092, August 2008.

[127] Rosalyn Moran, Dimitris A. Pinotsis, and Karl Friston. Neural masses and fields in dynamic causal modeling. *Frontiers in Computational Neuroscience*, 7, 2013.

[128] Georg Von Békésy. *Sensory Inhibition*. Princeton University Press, 1967.

[129] John Winn and Christopher M Bishop. Variational Message Passing. *Journal of Machine Learning Research*, page 34, 2005.

[130] Justin Dauwels. On Variational Message Passing on Factor Graphs. In *2007 IEEE International Symposium on Information Theory*, pages 2546–2550, Nice, June 2007. IEEE.

[131] Justin Dauwels, François Vialatte, Tomasz Rutkowski, and Andrzej Cichocki. Measuring Neural Synchrony by Message Passing. In *NIPS*, 2007.

[132] Dileep George. Belief Propagation and Wiring Length Optimization as Organizing Principles for Cortical Microcircuits. 2005.

[133] Gail A. Carpenter and Stephen Grossberg. A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37(1):54–115, January 1987.

[134] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, November 1998.

[135] Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025, November 1999.

[136] Wolfgang Maass. On the Computational Power of Winner-Take-All. *Neural Computation*, 12(11):2519–2535, November 2000.

[137] Quentin J. M. Huys, Neir Eshel, Elizabeth O'Nions, Luke Sheridan, Peter Dayan, and Jonathan P. Roiser. Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees. *PLoS Computational Biology*, 8(3):e1002410, March 2012.

[138] Stanislas Dehaene, Florent Meyniel, Catherine Wacongne, Liping Wang, and Christophe Pallier. The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron*, 88(1):2–19, October 2015.

[139] Jordi Fonollosa, Emre Neftci, and Mikhail Rabinovich. Learning of Chunking Sequences in Cognition and Behavior. *PLOS Computational Biology*, 11(11):e1004592, 19-Nov-2015.

[140] Masahiko Haruno, Daniel Wolpert, and Mitsuo Kawato. *Hierarchical MOSAIC for Movement Generation*. 2003.

[141] Emanuel Todorov. General duality between optimal control and estimation. *2008 47th IEEE Conference on Decision and Control*, pages 4286–4292, 2008.

[142] Bart van den Broek, Wim Wiegerinck, and Bert Kappen. Risk sensitive path integral control. *UAI*, 2010.

[143] W. H. Fleming and S. J. Sheu. Risk-sensitive control and an optimal investment model II. *The Annals of Applied Probability*, 12(2):730–767, May 2002.

[144] Andrew Barto, Marco Mirolli, and Gianluca Baldassarre. Novelty or Surprise? *Frontiers in Psychology*, 4, 2013.

[145] Emil Kauder. Genesis of the Marginal Utility Theory: From Aristotle to the End of the Eighteenth Century. *The Economic Journal*, 63(251):638–650, September 1953.

[146] Jürgen Schmidhuber. Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247, September 2010.

[147] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. Springer-Verlag, New York, 2 edition, 1985.

[148] Jürgen Schmidhuber. Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science*, 18(2):173–187, June 2006.

[149] Pierre-Yves Oudeyer and Frederic Kaplan. What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, 1, 2009.

[150] Edward Deci and Richard M. Ryan. *Intrinsic Motivation and Self-Determination in Human Behavior*. Perspectives in Social Psychology. Springer US, 1985.

[151] Ronald A. Howard. Information Value Theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1):22–26, August 1966.

[152] Laurent Itti and Pierre Baldi. Bayesian surprise attracts human attention. *Vision research*, 49(10):1295–1306, May 2009.

[153] Philipp Schwartenbeck, Thomas FitzGerald, Raymond J. Dolan, and Karl Friston. Exploration, novelty, surprise, and free energy minimization. *Frontiers in Psychology*, 4, October 2013.

[154] Yi Sun, Faustino Gomez, and Juergen Schmidhuber. Planning to Be Surprised: Optimal Bayesian Exploration in Dynamic Environments. *arXiv:1103.5708 [cs, stat]*, March 2011.

[155] H B Barlow. Inductive Inference, Coding, Perception, and Language. *Perception*, 3(2):123–134, June 1974.

[156] D. V. Lindley. On a Measure of the Information Provided by an Experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956.

[157] David J. C. MacKay. Information-Based Objective Functions for Active Data Selection. *Neural Computation*, 4(4):590–604, July 1992.

[158] Hagai Attias. Planning by Probabilistic Inference. In *9th Int. Workshop on Artificial Intelligence and Statistics*, page 8, 2003.

[159] Matthew Botvinick and Marc Toussaint. Planning as inference. *Trends in Cognitive Sciences*, 16(10):485–488, October 2012.

[160] Daniel Kahneman and Amos Tversky. *Prospect Theory: An Analysis of Decision under Risk*. Decision, Probability, and Utility: Selected Readings. Cambridge University Press, New York, NY, US, 1988.

[161] J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Theory of Games and Economic Behavior. Princeton University Press, Princeton, NJ, US, 1944.

[162] E. T. Jaynes. Information Theory and Statistical Mechanics. *Physical Review*, 106(4):620–630, May 1957.

[163] Baris Kurt. Kullback-Leibler Divergence Between Two Dirichlet (and Beta) Distributions, 2013.

[164] W.D. Penny. KL-divergence of Normal, Gamma, Dirichlet and Wishart densitites. 2001.

[165] Thomas Parr and Karl J. Friston. The Anatomy of Inference: Generative Models and Brain Structure. *Frontiers in Computational Neuroscience*, 12, 2018.

[166] Marjan Jahanshahi, Ignacio Obeso, John C. Rothwell, and José A. Obeso. A fronto-striato-subthalamic-pallidal network for goal-directed and habitual inhibition. *Nature Reviews. Neuroscience*, 16(12):719–732, December 2015.

[167] G. S. Berns and T. J. Sejnowski. How the Basal Ganglia Make Decisions. In A. R. Damasio, H. Damasio, and Y. Christen, editors, *Neurobiology of Decision-Making*, Research and Perspectives in Neurosciences, pages 101–113. Springer Berlin Heidelberg, Berlin, Heidelberg, 1996.

[168] Long Ding and Joshua I. Gold. The basal ganglia's contributions to perceptual decision-making. *Neuron*, 79(4):640–649, August 2013.

[169] Suzanne N. Haber. The primate basal ganglia: Parallel and integrative networks. *Journal of Chemical Neuroanatomy*, 26(4):317–330, December 2003.

[170] Florence Thibaut. Basal ganglia play a crucial role in decision making. *Dialogues in Clinical Neuroscience*, 18(1):3, March 2016.

[171] Karl Friston and Will Penny. Post hoc Bayesian model selection. *NeuroImage*, 56(4):2089–2099, June 2011.

[172] Ryan Smith, Philipp Schwartenbeck, Thomas Parr, and Karl J. Friston. An active inference model of concept learning. *bioRxiv*, page 633677, May 2019.

[173] Karl Friston, Thomas Parr, and Peter Zeidman. Bayesian model reduction. *arXiv:1805.07092 [stat]*, May 2018.

[174] Gerda Claeskens and Nils Lid Hjort. *Model Selection and Model Averaging*. Cambridge University Press, 2006.

[175] Klaas Enno Stephan, Will D. Penny, Jean Daunizeau, Rosalyn J. Moran, and Karl J. Friston. Bayesian model selection for group studies. *NeuroImage*, 46(4):1004–1017, July 2009.

[176] J.A. Hobson and K.J. Friston. Waking and dreaming consciousness: Neurobiological and functional considerations. *Progress in Neurobiology*, 98(1):82–98, July 2012.

[177] J. Allan Hobson, Charles C.-H. Hong, and Karl J. Friston. Virtual reality and consciousness inference in dreaming. *Frontiers in Psychology*, 5, 2014.

[178] G. Hinton, P Dayan, B. Frey, and R. Neal. The "wake-sleep" algorithm for unsupervised neural networks. *Science*, 268(5214):1158–1161, May 1995.

[179] Samuel J. Gershman and Yael Niv. Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology*, 20(2):251–256, April 2010.

[180] D. Gowanlock R. Tervo, Joshua B. Tenenbaum, and Samuel J. Gershman. Toward the neural implementation of structure learning. *Current Opinion in Neurobiology*, 37:99–105, April 2016.

[181] M. Berk Mirza, Rick A. Adams, Christoph Mathys, and Karl J. Friston. Human visual exploration reduces uncertainty about the sensed world. *PLOS ONE*, 13(1):e0190429, 05-Jan-2018.

[182] Harriet Brown, Rick A. Adams, Isabel Parees, Mark Edwards, and Karl Friston. Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14(4):411–427, November 2013.

[183] Rick A. Adams, Stewart Shipp, and Karl J. Friston. Predictions not commands: Active inference in the motor system. *Brain Structure & Function*, 218(3):611–643, May 2013.

[184] Karl Friston, Rick Adams, Laurent Perrinet, and Michael Breakspear. Perceptions as Hypotheses: Saccades as Experiments. *Frontiers in Psychology*, 3, 2012.

[185] Harriet Brown and Karl J. Friston. Free-Energy and Illusions: The Cornsweet Effect. *Frontiers in Psychology*, 3, 2012.

[186] Michael I. Jordan, Zoubin Ghahramani, and Lawrence K. Saul. *Hidden Markov Decision Trees*. 1997.

[187] David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. A learning algorithm for boltzmann machines. *Cognitive Science*, 9(1):147–169, January 1985.

[188] Ruslan Salakhutdinov and Geoffrey Hinton. An efficient learning procedure for deep Boltzmann machines. *Neural Computation*, 24(8):1967–2006, August 2012.

[189] K.J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *NeuroImage*, 19(4):1273–1302, August 2003.