Realizing Active Inference in Variational Message Passing: The Outcome-Blind Certainty Seeker

Théophile Champion
tmac3@kent.ac.uk
Marek Grzes
m.grzes@kent.ac.uk
University of Kent, School of Computing, Canterbury CT2 7NZ, U.K.

Howard Bowman
h.bowman@kent.ac.uk
University of Birmingham, School of Psychology, Birmingham B15 2TT, U.K.,
and University of Kent, School of Computing, Canterbury CT2 7NZ, U.K.

Active inference is a state-of-the-art framework in neuroscience that offers a unified theory of brain function. It is also proposed as a framework for planning in AI. Unfortunately, the complex mathematics required to create new models can impede application of active inference in neuroscience and AI research. This letter addresses this problem by providing a complete mathematical treatment of the active inference framework in discrete time and state spaces and the derivation of the update equations for any new model. We leverage the theoretical connection between active inference and variational message passing as described by John Winn and Christopher M. Bishop in 2005. Since variational message passing is a well-defined methodology for deriving Bayesian belief update equations, this letter opens the door to advanced generative models for active inference. We show that using a fully factorized variational distribution simplifies the expected free energy, which furnishes priors over policies so that agents seek unambiguous states. Finally, we consider future extensions that support deep tree searches for sequential policy optimization based on structure learning and belief propagation.

1 Introduction

The free energy principle aims to provide a unified theory of the brain based on Bayesian probability theory (Friston, 2010; Buckley, Kim, McGregor, & Seth, 2017). It takes root in Helmholtz’s argument that observations are produced by hidden causes that must be inferred and the predictive coding formulation that argues that inference and learning emerge from the reduction of the error between predicted and actual observations. Active inference
extends predictive coding to consider generative models of actions (Friston et al., 2016; Da Costa, Parr et al., 2020).

In brief, active inference is a probabilistic framework that describes how agents should act in their environment. It starts with the definition of a generative (probabilistic) model that encodes the agent’s beliefs about its environment. However, active inference does not rely on one particular generative model; instead it refers to a class of generative models that consider the impact of their actions in their environment. Active inference also relies on learning and inference to estimate the most likely states of the world and values of the model parameters. However, the concept behind active inference does not depend on a particular inference method, which means that both variational inference (Fox & Roberts, 2012) and Monte Carlo Markov chains (Fountas, Sajid, Mediano, & Friston, 2020) in principle can be used.

Active inference has been successfully applied in neuroscience to explain a wide range of brain phenomena such as habit formation (Friston et al., 2016), Bayesian surprise (Itti & Baldi, 2009), curiosity (Schwartenbeck et al., 2019), and dopaminergic discharges (FitzGerald, Dolan, & Friston, 2015). Active inference is also a form of planning as inference (Botvinick & Toussaint, 2012) consistent with Occam’s razor (Blumer, Ehrenfeucht, Haussler, & Warmuth, 1987) and can be seen as a generalization of reinforcement learning (van Hasselt, Guez, & Silver, 2015; Lample & Chaplot, 2016) and Kullback-Leibler control (Rawlik, Toussaint, & Vijayakumar, 2013). This framework has also been used to ground active vision (Ognibene & Baldassare, 2015; Heins et al., 2020; Van de Maele, Verbelen, Çatal, De Boom, & Dhoedt, 2021; Mirza, Adams, Mathys, & Friston, 2016, 2018) within a strong theoretical framework.

This letter focuses on active inference using variational (aka approximate Bayesian) inference and highlights its connection to variational message passing (Winn & Bishop, 2005). This ubiquitous message passing algorithm builds on the variational inference literature by leveraging the structure of the generative model to split the update equations into messages. Those messages transmit information about the new observations and, by summing those messages, it is possible to compute the posterior distribution over the parameters. The decomposition of the updates into messages formalizes the modularity of the method while remaining biologically plausible (Friston, Parr, & de Vries, 2017). Indeed, a key question in machine learning and computational neuroscience is how to identify compositional models, an issue that was identified early in the development of connectionism (Bowman & Li, 2011; Fodor & Pylyshyn, 1988). The central requirement being that higher-order representations (whether syntactic, semantic, perceptual, or something else) can be constructed by plugging together lower-order representations in such a way that the meanings of lower-order representations do not change (e.g., the “Jane” in “Jane loves John” is the
same “Jane” as in “John loves Jane”). It may be that the structural modularity provided by message passing implementations of Bayesian networks enable the compositionality of representations. According to modern trends, we use the formalism of Forney factor graphs (Forney, 2001) to represent the updates as messages sent along the graph edges.

Forney factor graphs are graphical representations used to realize generative models. They comprise two kinds of round nodes that represent the observations and the latent variables of the model. If the notion of observations can be understood as the data available to the model, the notion of latent variables is a bit more abstract. As an example, let us consider the MNIST data set (LeCun & Cortes, 2010) composed of images of handwritten digits. In this example, the pixels are observations made by the model, and latent variables could be any variables encoding the digit being represented, such as its orientation or size. The last type of nodes—square nodes—represent the dependency between observed and latent variables. In other words, how does the digit being represented generate the pixels?

The first goal of this letter is to provide readers with a full intuition of the mathematics underlying active inference and variational message passing. Then we show how to derive the update equations for any new generative models. The hope is to facilitate the development of new models that could, for example, play Atari games or model new brain mechanisms. Finally, we use our new generative model to prove that the update equations of active inference can be understood as variational message passing. This formal proof complements previous work that frames active inference as belief propagation (Friston, Parr et al., 2017) and enables us to create an automatic and modular implementation of active inference (van de Laar & de Vries, 2019a; Cox, van de Laar, & de Vries, 2019). This message passing formulation has particular consequences for the expected free energy, which is effectively reduced by the change, resulting in an agent that seeks certainty, without any concern for outcomes, whether preferred or not. We argue that the resulting behavior may have similarities to repetitive actions (sometimes called stimming) that are common, for example, in autism (Gabriels, Cuccaro, Hill, Ivers, & Goldson, 2005).

Section 2 describes the problem used to present the (classic) model widely used in the active inference literature. Sections 3 and 4 introduce variational inference and Forney factor graphs, respectively. Section 5 presents active inference as a decision theory based on the Bayesian view of probability, followed by section 6 that introduces the notion of variational message passing. Section 7 formulates active inference as variational message passing under a fully factorized approximate posterior (i.e., variational distribution) and explains the implications of this approximation for the expected free energy that underwrites policy selection. Before starting the next section, readers new to the active inference literature might want to read
appendix D, which uses Bayes theorem to present the simplest generative model sufficient for active inference.

## Problem Statement

Active inference crops up in many areas that require an agent to interact with its environment. Throughout this letter, the explanations are based on an agent named Bob, whose goal is to solve the food problem presented in section 2.2. But before we investigate this problem, let us have a look at how to simulate the interaction between Bob and his environment.

### 2.1 Simulating Active Inference

Most living beings are able to sense their environment through sensory inputs and process this sensory information to act in the world. For example, carnivorous flowers use tiny trigger hairs on their leaves to detect flies (sensing). When those hairs are stimulated, the ion concentrations in the leaves increase (processing), resulting in an electrical current that closes the leaf trapping the fly (acting). Similarly, humans gather sensory information through their five senses (sensing), process this information to understand their environment (processing), and, finally, make use of this understanding to act with intelligence (acting).

Sensing, processing, and acting correspond to the three steps of the action-perception cycle. This cycle conveniently casts active inference as an infinite loop (van de Laar & de Vries, 2019b). Each iteration begins by sampling the environment to obtain an observation, which is provided to the agent. Then the observation is used to perform inference (and learning) that produce a higher level of understanding; for example, an image might be mapped to a representation of the objects that it contains. And finally, this representation is exploited when acting to prepare your diner, drive your kids to school, or solve your favorite math problem.

### 2.2 The Food Problem

This section is concerned with the food problem initially proposed by Solopchuk (2018). This problem concerns an agent, named Bob, striving to survive. To produce the energy needed by his body, Bob needs to ingest nutriments. During periods of starvation, Bob’s stomach produces a hormone called ghrelin. This hormone travels to the brain through the blood and reaches a part of the brain named the hippocampus. This area has been shown to monitor the level of ghrelin in the blood (Kojima & Kangawa, 2005). At the moment ghrelin reaches the hippocampus, Bob’s brain can estimate the content of his stomach. This information can then be exploited to choose between eating and sleeping. However, the best action depends on the outcomes that Bob wants to witness in the future. This letter assumes that mother nature has kindly set Bob’s preferences to be biased toward the sensation of feeling fed (i.e., Bob enjoys observing low levels of ghrelin in his blood), which is arguably a favorable trait under a Darwinism view of evolution. Figure 1 summarizes the food problem.
3 Variational Inference

In Bayesian statistics, one assumes a prior distribution over latent (aka hidden) variables that represent the process generating the data. When more data are being collected, new observations bring information, allowing us to update our prior knowledge. The process of computing the most likely values of the hidden variables is called inference. A simple inference method is to use Bayes’ theorem to obtain the posterior probability distribution over the latent variable(s) of the model:

$$P(S|O) \propto \frac{P(O|S)P(S)}{P(O)} = \frac{P(O|S)P(S)}{\sum_S P(O|S)P(S)}$$

Since Bayes’ theorem is a corollary of the product rule of probability and no approximation is needed, it belongs to the field of exact inference. However, the computation of the evidence requires the marginalization over all hidden variables, which makes it intractable for all but the simplest models.
To address this intractability, one can turn to approximate or sampling-based methods. Variational inference belongs to the former and relies on an assumption of independence. As will be explained in section 6.1, the idea behind variational inference is to use a distribution $Q(S)$ to approximate the true posterior $P(S|O)$. This can be accomplished by minimizing the Kullback-Leibler (KL) divergence between some approximate and the true posterior:

$$D_{\text{KL}}[Q(S) \| P(S|O)].$$

Minimizing this KL divergence is impossible because the true posterior $P(S|O)$ is unknown. Fortunately, however, it is equivalent to minimizing the variational free energy $F$, known in machine learning as the negative evidence lower bound (ELBO). The variational free energy is defined as the Kullback-Leibler divergence between the variational distribution $Q(S)$ and the generative model $P(O, S)$:

$$F = D_{\text{KL}}[Q(S) \| P(O, S)] = -\text{ELBO}$$

$$= D_{\text{KL}}[Q(S) \| P(S|O)] + \ln P(O).$$

The variational distribution $Q(S)$ is used to approximate the true posterior $P(S|O)$. In addition to the introduction of this approximate posterior, the mean-field approximation makes the computation tractable by assuming that all latent variables are independent:

$$Q(S) = \prod_i Q_i(S_i),$$

where $Q_i(S_i)$ is the distribution over the $i$th hidden state of the model and $Q(S)$ is the joint distribution over all latent variables. This assumption of independence constrains the expressiveness of the variational distribution but allows the derivation of update equations, which can be evaluated efficiently.

At this point, an analogy might be useful to furnish an intuitive understanding of variational inference. Imagine you drop some coffee on a table, producing a stain with a complex shape. To compute the area of the stain, it might be useful to first assume an elliptic shape for the stain. However, since the stain is not actually elliptic, the solution will be only an approximation. In this analogy, the stain is the true posterior, and the ellipse is the approximate posterior.

This analogy should help with understanding Figure 2, which illustrates the kinds of results obtained by variational methods. As will be demonstrated in section 6.2, it is possible to prove (Fox & Roberts, 2012) that minimizing the variational free energy $F$ with respect to $Q_i(S_i)$ can be performed...
Realizing Active Inference in Variational Message Passing

Figure 2: This figure illustrates the kind of result obtained using variational inference. The true posterior drawn in red has a complex shape and is approximated by the variational distribution drawn in blue. The gray area depicts the error made when using the variational distribution to approximate the true posterior.

by iterating one of the following update equations:

\[
\ln Q_k(S_k) \leftarrow \ln Q^*_k(S_k) = (\ln P(O, S))_{\sim Q_k},
\]

\[
\Leftrightarrow Q_k(S_k) \leftarrow Q^*_k(S_k) = \frac{1}{Z} \exp (\ln P(O, S))_{\sim Q_k},
\]

where \( Q^*_k(S_k) \) is the optimal posterior, \( Z \) is a normalization constant, and \( (\cdot)_{\sim Q_k} \) is the expectation over all factors but \( Q_k \). Importantly, it is the coupling of the above update equations (i.e., one update per hidden variable \( S_k \)) that justifies the iteration of the updates until convergence to the free energy minimum.

4 Forney Factor Graphs

Typically, generative models are represented graphically using a graphical model (Koller & Friedman, 2009) or Forney factor graph (Forney, 2001). This section focuses on the latter representation introduced by David Forney in 2001, which uses three kinds of nodes. The nodes representing hidden and observed variables are depicted by white and gray circles, respectively, and factors are represented using white squares, which are linked to variable nodes by arrows or lines. Arrows are used to connect factors to their target variable, while lines link factors to their predictors. Figure 3 shows an example of a Forney factor graph corresponding to the following generative model:

\[
P(O, S) = P_O(O|S)P_S(S).
\]

Generally factor graphs describe only the model’s structure, in terms of the variables and their dependencies, but not the individual factors. For
Figure 3: This figure illustrates the Forney factor graph corresponding to the following generative model: $P(O, S) = P_S(O|S)P_S(S)$. The hidden state is represented by a white circle with the variable’s name at the center, and the observed variable is depicted similarly but with a gray background. The factors of the generative model are represented by squares with a white background and the factor’s name at the center. Finally, arrows connect the factors to their target variable, and lines link each factor to its predictor variables.

For example, the definitions of $P_O$ and $P_S$ are not given by Figure 3, and additional information is required; for example, $P_S(S) = \mathcal{N}(S; \mu, \sigma)$ specifies $P_S$ as a Gaussian distribution.

Initially, variables could connect to only a limited number of factors. However, a special kind of factor, called an equality node, dissolves this limitation. Purists tend to represent all equality nodes, while others make them implicit by allowing the variables to connect to an arbitrary number of factors. For clarity, this letter keeps equality nodes implicit.

Finally, factors, along with hidden and observed variables, are sometimes called constraint, state, and symbol, respectively. As explained by Yedidia (2011), those two terminologies refer to two views on Forney factor graphs, where factors encode probabilities and constraints encode costs. Infinite costs represent hard constraints, while finite costs encode soft constraints. Here, hard constraints define which configurations of the state space are forbidden (i.e., have a probability of zero), and soft constraints encode preferences over the state configurations (i.e., the higher the cost the smaller the state probability). This reveals an interesting link between Bayesian statistics and symbolic artificial intelligence and prompts the question of whether Bayesian statistics can be regarded as a generalization of symbolic artificial intelligence. For example, one could start by framing the problem of constraint satisfaction as an inference process on a Forney factor graph that encodes the problem constraints.
Realizing Active Inference in Variational Message Passing

5 Active Inference

So far, we have discussed variational inference and Forney factor graphs. We now present the intuition behind the various equations that comprise the active inference framework. We will be working with the food problem introduced in section 2.

5.1 Generative Model. We begin by presenting the generative model introduced by Friston et al. (2013). Instead of presenting the full generative model at once, we build this model progressively. This should help readers to understand both the model and its corresponding Forney factor graph.

5.1.1 The D Vector. As we shall see shortly, the full generative model represents the world as a sequence of hidden states, and those states generate the observations made by the agent. Those states are arranged chronologically using the index \( \tau \) that runs from the initial state \( (S_0) \) to the state of the last time step \( (S_T) \). This section focuses on the initial state, whose distribution is a categorical, defined as

\[
P_{S_0}(S_0 | D) = \text{Cat}(S_0; D),
\]

where \( D \) is a vector containing the parameters of the categorical distribution. In addition to the categorical distribution, the model assumes a Dirichlet prior over the parameters \( D \), leading to

\[
P_D(D) = \text{Dir}(D; d).
\]

In this context, the parameters \( d \) of the Dirichlet distribution are called hyperparameters, because they control the distribution of the parameters \( D \). Figure 4 summarizes this part of the model by presenting an example of the vector \( D \) and the Forney factor graph corresponding to the two distributions constituting Bob’s generative model.

5.1.2 The A Matrix. We have already mentioned that the probability of an observation (aka outcome), such as feeling hungry, depends on the value of the hidden state: whether Bob’s stomach is full or empty. This dependency is represented by a conditional distribution, such that the likelihood of an observation, given a particular value of the hidden states, is defined by a categorical distribution as follows:

\[
P_{O_\tau}(O_\tau | S_\tau = j, A) = \text{Cat}(O_\tau; A_{\cdot j}),
\]

where the \( j \)th column of \( A \), denoted \( A_{\cdot j} \), contains the parameters of the categorical distribution encoding the probability of the outcomes given that
Figure 4: The vector $D$ that defines Bob’s beliefs about the initial hidden state, and the Forney factor graph corresponding to equations 5.1 and 5.2. Since the probability of $S_0$ being full is higher than the probability of it being empty, Bob thinks that at the beginning of each trial, his stomach is more likely to be full than empty.

$S_\tau = j$. Additionally, we can rewrite the above equation more concisely by letting $S_\tau$ be a one hot vector, whose $j$th element is equal to one, such that

$$P(O_\tau | S_\tau, A) = \text{Cat}(O_\tau; A S_\tau),$$

where because $S_\tau$ is a one hot vector, the multiplication of $A$ and $S_\tau$ selects the $j$th column of $A$. Similar to the treatment of the vector $D$, a prior over the columns of $A$ is used. To ensure the conjugacy between the distributions of the model, a Dirichlet prior is used for each column. The probability of the overall matrix is then given by the following product of Dirichlet distributions:

$$P_A(A) = \prod_i \text{Dir}(A_{i}; a_i),$$

where $a$ is a matrix containing the parameters of the Dirichlet distributions (each column of $a$ contains the parameters of one Dirichlet distribution). Note that because each column of the matrix $A$ is a categorical distribution, the conjugate prior of each column is a Dirichlet distribution. Assuming independence of the columns of $A$, the conjugate prior of the entire matrix $A$ is a product of Dirichlet distributions. Importantly, the prior over $A$ is not a Dirichlet distribution whose parameters are obtained by concatenation of the columns of $A$. Indeed, if we sample from such a (concatenated) prior, then the elements of the entire matrix will sum up to one but the columns would not. This is problematic because each column of $A$ is supposed to be a categorical distribution that sum up to one. We conclude this section...
Figure 5: The matrix $A$ that defines how the hidden states generate the observations. In our example with Bob, this matrix defines the probability of Bob feeling hungry or fed while his stomach is full or empty. Furthermore, the new version of the generative model is shown on the right.

with Figure 5, which illustrates the likely matrix $A$, along with the resulting version of the generative model for Bob’s problem.

5.1.3 The $B$ Matrices. Now that readers are familiar with the definition of the likelihood matrix $A$, we focus on the temporal transitions between any pair of successive states. Those transitions are modeled similar to the matrix $A$ that concerns the generation of observations from hidden states. However, here we are concerned with the transition matrices that maps from states at one time point to the next. Crucially, there are as many of these matrices as the number of allowable actions on the state in question. This follows from the idea that each action has the potential to modify Bob’s stomach differently; for example, eating is more likely to change Bob’s stomach from empty to full than sleeping would. Accordingly, the transition between two consecutive hidden states is defined by a set of matrices, called the transition or $B$ matrices, such that

$$P_{S_{t+1} | S_t = i, \pi = j, B} = \text{Cat}(S_{t+1}; B[U_i]_{i=1}^j),$$

where $\text{Cat}(\cdot)$ means equal by definition, $U \triangleq U^i$ is the action predicted at time step $t$ by the $j$th policy, and $B[U]$ is the matrix corresponding to the action $U$. Furthermore, active inference defines policies as action sequences (see the next section). By replacing the index $i$ by a one hot vector as in the previous section, equation 5.3 can be rewritten as

$$P_{S_{t+1} | S_t, \pi, B} = \text{Cat}(S_{t+1}; B[U]S_t).$$
The matrices $B$ that define the transition between any two consecutive hidden states. In the context of the food problem, these matrices encode the probability of transitioning from a full or empty stomach at time $\tau$ to a full or empty stomach at time $\tau + 1$.

A Dirichlet prior is assumed for each column of the transition matrices $B$, leading to the following prior

$$P(B) = \prod_{i,j} \text{Dir}(B[i,j]; b[i,j]),$$

where $b$ are the parameters of the Dirichlet distributions and $i$ and $j$ iterate over all possible actions and states, respectively. Finally, Figures 6 and 7 illustrate the matrices $B$ and the updated version of the generative model.
5.1.4 The Prior over Policies. We now consider the prior over the policy that was left undefined in Figure 7. But what do we exactly mean by “policies”? In active inference, a policy is a sequence of actions over time, \( \{U_t, \ldots, U_{T-1}\} \). As a consequence, even if the agent expects the environment to be in the same state at two different time steps, picking two different actions at those time steps is still possible. Therefore, an active inference agent can perform an epistemic action as long as there is some uncertainty to be reduced and then switch to exploitative behaviors. Note that this definition of policy is in opposition to most of the model-free reinforcement learning literature, where a policy is a mapping from states to actions. In particular, states in the context of model-free reinforcement learning are observed and therefore are closer to the notion of observations in active inference. Technically, active inference takes us out of the world of fixed state-action policies (where the same action is taken from each state) into the world of sequential policy optimization, where different actions can be taken from the same state—crucially, in a way that depends on (Bayesian) beliefs about hidden states.

The last ingredient required to obtain the prior over the policies is a notion of policy quality. In active inference, good policies are the ones that minimize the expected free energy—that is, the free energy expected in the future, which is defined as

\[
G(\pi) \approx \sum_{t=t+1}^{T} \left[ \sum_{\tau=1}^{\tau} D_{KL}[Q(O_{\tau} | \pi) \| P(O_{\tau})] + \mathbb{E}_{Q(S_{\tau} | \pi)}[H[P(O_{\tau} | S_{\tau})]] \right],
\]

where \( H[\cdot] \) is the Shannon entropy, \( G \) is a vector containing as many elements as the number of policies, and the \( i \)th element of \( G \) represents the quality of the \( i \)th policy. (Readers who are interested in the derivation of the expected free energy are referred to appendix C.) We should mention here that \( Q(O_{\tau} | \pi) \) and \( Q(S_{\tau} | \pi) \) are computed based on the result of the inference process of the previous action-perception cycle. Therefore, \( G \) can be regarded as a model parameter and is not represented as a random variable in the Forney factor graph. The definition and justification of the expected free energy are provided in appendix C and a recent paper by Millidge, Tschantz, and Buckley (2020). Also, the expected free energy arises naturally in mathematical treatments of the free energy principle when considering self-organization at nonequilibrium steady state (Friston, 2019; Parr, Da Costa, & Friston, 2020). At this point, we should take a moment to understand the intuition behind the expected free energy.

Let us begin with the second term of equation 5.4. For each value of the hidden state, \( P(O_t | S_t = i) \) is a categorical distribution whose parameters correspond to the \( i \)th column of \( A \). This distribution defines the probability
Next, we need to encode Bob’s preferences over future outcomes, which are called prior preferences. Formally, those preferences are defined as a categorical distribution whose parameters are stored in the vector $C$. Figure 8 illustrates this vector. It should be noted that those preferences define the goodness of future outcomes, and we shall come back to this when discussing the link between active inference and reinforcement learning (see appendix A).

To conclude, we need to consider the predicted or expected outcomes. One way to predict future outcomes would be to compute the marginal distribution over $O_t$ using, for example, the sum product algorithm (Kschischang, Frey, & Loeliger, 2001). However, this might be computationally expensive, so we will proceed with the following formula,

$$Q(O_t | \pi) = \sum_i P(O_t | S_t = i, A) Q(S_t = i | \pi) = A s^T,$$

where, as will be discussed in section 5.2, $Q(S_t | \pi) \approx \text{Cat}(S_t; s^T)$. This equation can be understood as a form of marginalization, where the approximate posterior $Q(S_t | \pi)$ is our most informed belief about the hidden states. Finally, the KL divergence between the expected outcomes and the prior preference is called risk (see appendix A for additional details). The risk
Figure 9: A distribution over the policies that gives high probability to policies fulfilling Bob’s preferences in the future. For example, the first policy, where Bob is constantly eating, has high probability, while the fourth policy, where Bob is constantly sleeping, has low probability. This is congruent with the notion that eating is more likely to make Bob feel fed than hungry, and similarly, sleeping is more likely to make Bob feel hungry than fed.

Part of expected free energy is simply the divergence between the expected outcomes and the preferred outcomes. It is this part of expected free energy that underwrites policies that lead to preferred outcomes under uncertainty. Minimizing expected free energy therefore minimizes risk (i.e., the divergence between anticipated and preferred outcomes) and ambiguity (i.e., the conditional uncertainty about outcomes, given the causes). The resulting prior over the policies is defined as

$$P_\pi(\pi | \gamma) = \sigma(-\gamma G),$$

where $\sigma(\cdot)$ is the softmax function, $G$ is the expected free energy, $\gamma$ determines the sensitivity of policy selection to the expected free energy of each policy, and the negative sign gives high probability to policies minimizing expected free energy. Importantly, the prior over policies is an empirical prior because the expected free energy depends on the observations, which means that it must be reevaluated each time a new observation is made by the agent. In other words, the prior over the policies is a Boltzmann distribution with $\gamma$ being the inverse temperature. Taking this view, small values for $\gamma$ means a high temperature and less precise prior beliefs about which policy should, or is, being pursued. Figure 9 shows an example of this distribution, and Figure 10 illustrates the current generative model.

5.1.5 The Prior over the Precision Parameter. We now turn to the last part of the generative model: the prior over the precision parameter $\gamma$. Importantly, this precision parameter has been associated with the neuromodulator dopamine through what is called the “precision hypothesis”
Figure 10: The Forney factor graph of the entire generative model of the sort presented by Friston et al. (2016). Section 5.1.1 described how the probability of the initial states is defined by the vector \( D \), and as discussed in section 5.1.2, the matrix \( A \) defines the probability of the observations given the hidden states. Section 5.1.3 explained that the \( B \) matrices define the transition between any successive pair of hidden states. This transition depends on the action performed by the agent, that is, on the policy \( \pi \). Furthermore, the prior over the policies has been chosen in section 5.1.4, such that policies minimizing the expected free energy are more probable. Finally, we see in section 5.1.5 that the precision parameter \( \gamma \) (which modulates the stochasticity of the agent behavior) is distributed according to a gamma distribution.

(FitzGerald et al., 2015). This association of dopamine and the precision parameter claims to unify two perspectives on the role of dopamine. The first frames dopamine as an error signal on predicted reward (Schultz, Dayan, & Montague, 1997) and uses the framework of TD learning. The second, called the incentive salience hypothesis, frames dopamine as “associating salience and attractiveness to visual, auditory, tactile, or olfactory stimuli” (Berridge, 2007).

But let us come back to the prior over the precision parameters \( \gamma \). In neurobiological treatments, this prior usually takes the form of a gamma distribution with a rate parameter \( \beta \) and a shape parameter fixed to one:

\[
P_\gamma(\gamma) = \Gamma(\gamma; 1, \beta).
\]
Realizing Active Inference in Variational Message Passing

Figure 11: Four gamma distributions where the values of the parameters have been changed. The graph on the right shows the kind of prior the model believes in by forcing the shape parameter to equal one.

The graph on the right of Figure 11 illustrates two variations of this prior for $\beta = 1$ and $\beta = 2$. Also, we should mention that a more flexible prior can be obtained by removing the constraint on the shape parameter (Friston et al., 2015), and the left side of Figure 11 illustrates this extension. However, in most artificial intelligence applications (that are not concerned with biological implementation or dopamine), $\gamma$ is usually assumed to be one. Mainly, this design choice is made for simplicity, even if in practice, forcing $\gamma$ to be one reduces the model flexibility, that is $\gamma$ can no longer be learned.

5.1.6 The Entire Generative Model. Throughout this section, we have assembled incrementally the generative model usually used in active inference, whose Forney factor graph is represented in Figure 10. The last step is to write down the equations that constitute its formal definition:

$$P(O_0, S_{0:T}, \pi, A, B, D, \gamma) = P(\pi | \gamma)P(\gamma)P(A)P(B)P(S_0)P(D)$$

$$\prod_{t=0}^{T} P(O_t | S_t, A) \prod_{t=1}^{T} P(S_t | S_{t-1}, B, \pi), \quad (5.5)$$

where:

- $P(\pi | \gamma) = \sigma(-\gamma G)$
- $P(\gamma) = \Gamma(\gamma; 1, \beta)$
- $P(A) = \prod_{i} \text{Dir}(A_i; a_i)$
- $P(B) = \prod_{i,j} \text{Dir}(B[i,j]; b[i,j])$
- $P(S_0 | D) = \text{Cat}(S_0; D)$
- $P(D) = \text{Dir}(D; d)$
- $P(O_t | S_t, A) = \text{Cat}(O_t; AS_t)$
- $P(S_t | S_{t-1}, B, \pi) = \text{Cat}(S_t; B[S_t]S_{t-1})$. 


Table 1: Generative Model Notation.

| Notation | Meaning |
|----------|---------|
| $T$ | The time horizon |
| $t$ | The current time steps |
| $\tau$ | An iterator over time step |
| $O_{0:t}$ | The sequence of observations between time step 0 and t |
| $S_{0:T}$ | The sequence of hidden states between time step 0 and T |
| $\pi$ | The policies |
| $U_m^{\tau}$ | The action or control state predicted by the $m$th policy at time step $\tau$ |
| $A$ | The matrix defining the likelihood mapping from the hidden states to the observations |
| $A_{i}$ | The $i$th column of the matrix $A$ |
| $B$ | The set of transition matrices defining the mappings between any two consecutive hidden states |
| $B[U]_{i}$ | The $i$th column of the transition matrix $B[U]$ corresponding to action $U$ |
| $D$ | The prior over the initial hidden states |
| $a, b, d$ | The parameters of the prior over $A, B$, and $D$ |
| $a_{i}$ | The $i$th column of the matrix $a$ |
| $b[U]_{i}$ | The $i$th column of the matrix $b[U]$ corresponding to action $U$ |
| $\gamma$ | The precision parameter related to neuromodulators such as dopamine |
| $\sigma(x)$ | The softmax function |
| $G$ | The expected free energy |
| $\Gamma(\gamma; \alpha, \beta)$ | Gamma distribution with shape and inverse scale parameters $\alpha$ and $\beta$ |
| $\text{Cat}(S_0; D)$ | Categorical distribution over $S_0$ with parameter $D$ |
| $\text{Dir}(D; d)$ | Dirichlet distribution |

Note that to keep the notation uncluttered, we have dropped the subscripts such that $P_{S_0}(S_0|D)$ becomes $P(S_0|D)$, $P_A(A)$ becomes $P(A)$ and so forth. Table 1 provides a complete description of the notation used to define the generative model.

5.2 Variational Distribution. We now turn to the definition of the variational distribution, which is used to approximate the true posterior during variational inference (aka approximate Bayesian inference), that is, $Q(x) \approx P(x|o)$, where $x$ and $o$ denote the hidden variables and the observations, respectively. Let us first recall that variational inference leverages independence between latent variables in what is known as a mean-field approximation. A structured approximation, often made in the active inference literature\(^1\) to simplify computations, is that all latent variables are independent except for the hidden states and the policy. This leads to the following

\(^1\) An instance where this general assumption is not made can be found in Parr, Dimitrije, Kiebel, and Friston (2019).
Realizing Active Inference in Variational Message Passing

Table 2: Variational Distribution Notation.

| Notation | Meaning |
|----------|---------|
| $s_i$   | The parameters of the posterior over $S_i$ for each policy (i.e., a vector) |
| $s_\tau$ | The parameters of the posterior over $S_\tau$ for all policies (i.e., a matrix) |
| $\pi$   | The parameters of the posterior over $\pi$ (i.e., a vector) |
| $a, b, d$ | The parameters of the posterior over $A, B, D$ (i.e., a matrix, a set of matrices, and a vector, respectively) |
| $\beta$ | The (inverse temperature) parameter of the posterior over $\gamma$ |

Once again, for compactness, the subscript will be dropped (e.g., $Q_{S_\tau}(S_\tau|\pi)$ will be replaced by $Q(S_\tau|\pi)$). Table 2 summarizes the notation used to define this variational distribution. It is much easier to understand this distribution by comparing it to the definition of the generative model in equation 5.5. Indeed, the distributions over $A, B,$ and $D$ remain Dirichlet distributions, and the distributions over $\gamma$ and $S_\tau$ remain gamma and categorical distributions, respectively. Only the distribution over $\pi$ changes from a Boltzmann to a categorical distribution. However, both the Boltzmann and the categorical are discrete distributions.

5.3 Variational Free Energy. Above, we have unpacked the generative model and variational distribution used in active inference. This section combines those two concepts to form the second cornerstone of the active inference framework, the variational free energy. Section 6.1 explains how the following equation can be derived from the Kullback-Leibler divergence between the variational distribution and the true posterior. However, this section explains the intuition behind the variational free energy, which is
defined as follows

\[
F = \mathbb{E}_Q [\ln Q(S_{0:T}, \pi, A, B, D, \gamma) - \ln P(O_{0:t}, S_{0:T}, \pi, A, B, D, \gamma)]
\]

\[
= D_{KL} \left[ Q(x) \Big|\Big| P(x|\omega) \right] - \ln P(\omega),
\]

(5.7)

where \( x = \{S_{0:T}, \pi, A, B, D, \gamma\} \) refers to the model’s hidden variables, and \( \omega = \{O_{0:t}\} \) refers to the sequence of observations made by the agent. Equation 5.7 highlights some important properties of the variational free energy. Indeed, the relative entropy (aka KL divergence) ensures that the variational distribution \( Q(x) \) tends to get closer to the true posterior \( P(x|\omega) \) as the free energy is reduced. Furthermore, it shows that the variational free energy is an upper bound on the negative log evidence because the relative entropy cannot be negative. Also, if the variational distribution is equal to the true posterior, then the variational free energy is equal to the (-ve) log evidence. The variational free energy can also be rearranged as

\[
F = D_{KL} \left[ Q(x) \Big|\Big| P(x) \right] - \mathbb{E}_Q [\ln P(\omega|x)],
\]

(5.8)

showing the trade-off between complexity and accuracy. The complexity penalizes the divergence of the posterior \( Q(x) \) from the prior \( P(x) \). The accuracy scores how likely the observations are given the generative model and current belief of the hidden states. Interestingly, in opposition to the Akaike information criterion (AIC) and Bayesian information criterion (BIC), the complexity does not depend on the number of parameters. Consequently, a model with a lot of parameters but does not vary from the prior will have zero complexity, and a model with a small number of parameters that moves away a lot from the prior will have a large complexity. Taking this view, a model is complex whenever the knowledge encoded by the prior fails to explain the observed data accurately. In other words, complexity scores the degree of belief updating that moves posterior beliefs away from prior beliefs to provide an accurate account of any observations.

Comparison of the expression for expected free energy and variational free energy reveals an intimate relationship. One can see that the risk is the expected complexity, while ambiguity is expected inaccuracy. These expectations are under the posterior predictive beliefs about outcomes in the future under the policy in question. This is why \( G \) is called expected free energy.

### 5.4 Update Equations

All the update equations presented below come from the minimization of the variational free energy. This section presents the intuition behind those updates using the notations summarized in
Realizing Active Inference in Variational Message Passing

Table 3: Update Equations Notation

| Notation | Meaning |
|----------|---------|
| $A \otimes B = AB^T, A \cdot B = A^T B$ | outer and inner products |
| $[a, b]$ | all the natural numbers between $a$ and $b$ |
| $\Omega_a = \{(k, r) : U_{t-1}^k = u, r \in [1, T]\}$ | all $(k, r)$ such that the $r$th policy predicts action $u$ at time $t = k$ |
| $s_t = s_t^\pi \cdot \pi$ | the expected state at time $t$ |
| $\langle f(X) \rangle_{P_X} \triangleq \mathbb{E}_{P_X}[f(X)]$ | the expectation of $f(X)$ over $P_X$ |
| $\psi(x)$ | the digamma function used to compute analytical solutions, e.g., for $\langle \ln D_1 \rangle_{Q_0}$ |
| $D_t = \langle \ln D_t \rangle_{Q_0} = \psi(d_t) - \psi(\sum d_t)$ | the expected logarithm of $D$ |
| $A_{ij} = \langle \ln A_{ij} \rangle_{Q_0} = \psi(a_{ij}) - \psi(\sum a_{ij})$ | the expected logarithm of $A$ |
| $B[u]_{ij} = \langle \ln B[u]_{ij} \rangle_{Q_0} = \psi(b[u]_{ij}) - \psi(\sum b[u]_{ij})$ | the expected logarithm of $B$ |

Table 3. We start with the optimal updates of $A$, $B$, and $D$, which are given by

$$Q^*(D) = \text{Dir}(D; d) \quad \text{where} \quad d = d + s_0.$$  \hspace{1cm} (5.9)

$$Q^*(A) = \prod_i \text{Dir}(A_{ij}, a_{ij}) \quad \text{where} \quad a = a + \sum_{r=0}^{t} o_r \otimes s_r.$$  \hspace{1cm} (5.10)

$$Q^*(B) = \prod_{u,i} \text{Dir}(B[u]_{ij}, b[u]_{ij}) \quad \text{where} \quad b[u] = b[u] + \sum_{(k,r) \in \Omega_u} s_r^k \otimes s_{r-1}^k \pi_k.$$  \hspace{1cm} (5.11)

Looking at the above equations, these updates can be understood as counting the number of times an event appears. For example, the update of $A$ counts the number of times a pair of states-observations has been observed. Taking this view, $a$ is the pseudo-count of previously occurring states-observations pairs, and $o_r \otimes s_r$ takes into account the new observations. Similarly, the update of the $B$ and $D$ matrices, respectively, counts how many times the state transitions and initial states have been observed. Additionally, the updates of the hidden states are

$$Q^*(S_0 | \pi) = \sigma \left( D + I(0 \leq t) o_0 \cdot A + s_t^U \cdot B[U^T] \right).$$  \hspace{1cm} (5.12)

$$Q^*(S_t | \pi) = \sigma \left( B[U^T] s_{t-1}^T + I(t \leq l) o_r \cdot A + s_{t+1}^U \cdot B[U^T] \right).$$  \hspace{1cm} (5.13)

$$Q^*(S_T | \pi) = \sigma \left( B[U^T] s_{T-1}^T + I(T \leq l) o_T \cdot A \right).$$  \hspace{1cm} (5.14)
where $t$ can be thought of as a global variable referring to the present time point and $I(\cdot)$ is an indicator function that equals one if the condition is true and zero otherwise. A closer look at these updates reveals that the hidden states are updated by gathering information from the past, the future, and the likelihood mapping. In equation 5.12, the information from the past is replaced by some information from the prior over the initial state, and in equation 5.14, the information from the future disappears because we have reached the limits of the time horizon (i.e., $\tau = T$). Similarly, in equations 5.13 and 5.14, the indicator function ensures that there is no information from the likelihood mapping after the current time step $t$ because no observations are available. (For additional information about the above updates, see sections 7.7 and 7.8 as well as appendix G.) Interestingly, Parr and Friston (2018) proposed a model in which future observations are latent variables, and in this case, information will be sent along the edges connecting future states and future observations. Finally, the update of $\gamma$ and $\pi$ takes the following form,

$$Q^*(\gamma) = \Gamma\left(\gamma; 1, \beta + G \cdot (\pi - \pi_0)\right),$$

$$Q^*(\pi) = \sigma\left(-\frac{1}{\beta}G - F\right),$$

where $\pi_0 = \sigma(-\gamma \cdot G)$, $\sigma(\cdot)$ is the softmax function, and $F$ is a vector whose $\pi$th element is defined as

$$F_\pi = s_0^\pi \cdot (\ln s_0^\pi - D) + \sum_{t=1}^{T} s_t^\pi \cdot (\ln s_t^\pi - B[U]s_{t-1}) - \sum_{\tau=0}^{l} o_{\tau} \cdot \bar{A}s_{\tau}.$$

Section 7 derives update equations similar to those above that can be decomposed as a sum of messages coming from the parent, children, and co-parents of each node.

5.5 Action Selection. This section focuses on the various strategies available to pick the next action(s) that the agent will then perform. In active inference, the action selection process is performed after iteration of the update equations. Indeed, according to the action-perception cycle presented in section 2, the agent first minimizes the variational free energy and then acts in its environment. The first strategy entails summing the posterior evidence for the policies predicting each action and executing the action with the highest sum of posterior evidence:

$$u_t^* = \arg\max_u \sum_{m=1}^{||\pi||} \delta_{u, U^m} Q(\pi = m).$$
where \(|\pi|\) is the number of policies, \(U_m^t\) is the action predicted at the current time step by the policy \(\pi\), and \(\delta_{u,U_m^t}\) is an indicator function that equals one if \(u = U_m^t\) and zero otherwise. Since the model knows the posterior over the policies (i.e., sequences of actions) another strategy is to simply sample an entire policy (e.g., a sequence of actions) without recomputing the posterior at each time step, that is, Bob selects a policy, closes his eyes, and performs the sequence of actions entailed by that multistep policy. In the case of single-step policies, this is equivalent to the first strategy. This leads to a trade-off between computational time and the quality of the actions selected. Indeed, the more actions selected at once, the less computational time required, but the less informed those actions will be.

Another strategy used in planning is a Monte Carlo tree search (Browne et al., 2012). The best known example of Monte Carlo tree search is probably the victory of AlphaGo against Lee Sedol (the go world champion) in 2016 (Silver et al., 2016). Interestingly, this method has been used recently with an active inference agent (Fountas et al., 2020). The simplest version of this algorithm starts with an empty tree, that is, a single node representing the current state. Then the root node is expanded such that the states that are reachable from the current state become its children. Those children are linked to the root node by edges representing the actions leading to those states. Afterward, simulations of the environment are run to evaluate how good those new child states are. In the context of reinforcement learning, the goodness of the states corresponds to whether rewarding terminal states are reached during the simulations. Similarly, in the context of active inference, the expected free energy scores the goodness of outcomes. Finally, the reward or EFE is backpropagated upward in the tree. Iterating this four-step process (selection, expansion, simulation, and backpropagation) furnishes a posterior over the best action to perform next.

6 Variational Message Passing

In the previous sections, our focus was on explaining the intuition behind active inference. This section is more technical. We begin with the KL divergence between the variational distribution \(Q(x)\) and the true posterior \(P(x|o)\), which underwrites the minimization of the variational free energy. Then we derive two update equations well known from the Bayesian statistics community. The first explains how the approximate posterior can be computed using variational inference. And the second reveals that the optimal posterior can be thought of as a sum of messages. Finally, the message-based equation is specialized for the class of exponential conjugate models that we use to describe the method of Winn and Bishop (2005) as a five-step process. In this section, we use a few properties that are summarized in appendix B.
6.1 Justification of the Variational Free Energy. As mentioned in section 3, the computation of the true posterior using Bayes’ theorem quickly becomes intractable as the number of hidden states increases. The variational free energy (VFE) or, equivalently, the negative evidence lower bound (-ELBO), aims to solve this intractability problem by approximating the true posterior with another distribution: the variational distribution. To justify the use of the variational free energy, we first note that the following expression can be obtained from the product rule:

\[ P(x|o) = P(o, x) / P(o). \] (6.1)

Since the KL divergence measures the distance between two distributions, we can minimize the KL divergence between the variational distribution and the true posterior. And this will keep the variational distribution close to the true posterior. Starting with this KL divergence, and substituting equation 6.1 within it, we obtain

\[
D_{KL}[Q(x)\|P(x|o)] = \mathbb{E}_{Q(x)}[\ln Q(x) - \ln P(x, o)]
\]

where the expectation over the log evidence can be dropped due to the lack of a dependence of \(\ln P(o)\) on \(Q(x)\). Because the log evidence does not depend on the latent variables, it can be safely ignored during the minimization process. In other words, minimizing the variational free energy is equivalent to minimizing the KL divergence between the variational distribution and the true posterior and ensuring that the variational distribution is a good approximation of the true posterior.

6.2 Variational Inference Updates. As we have just noted, variational methods rely on the minimization of the variational free energy or, equivalently, the maximization of an evidence lower bound. So let us start with the former:

\[
D_{KL}[Q(x)\|P(o, x)] = \mathbb{E}_{Q(x)}[\ln Q(x) - \ln P(x, o)].
\]

Using the mean-field assumption, \(Q(x) = \prod_i Q_i(x_i)\), the log property, and the linearity of expectation. The above equation can be rewritten as

\[
D_{KL}[Q(x)\|P(o, x)] = \mathbb{E}_{Q(x)}[\ln Q(x) - \ln P(x, o)]
\]
Realizing Active Inference in Variational Message Passing

Note that $\ln Q_k(x_k)$ is a constant with regard to all factors but $Q_k(x_k)$, and $\ln \prod_{j\neq k} Q_j(x_j)$ is a constant with regard to $Q_k(x_k)$. Using the expectation of a constant, the above equation can be rewritten as

$$D_{KL} \left[ Q(x) \left| \left| P(o, x) \right. \right. \right]$$

$$= E_{Q,(x_k)}[\ln Q_k(x_k)] + E_{-Q_k(x_k)}[\ln \prod_{j\neq k} Q_j(x_j)] - E_{Q_k(x_k)}[\ln P(x, o)],$$

where $E_{-Q_k(x_k)}[\cdot]$ is the expectation over all factors but $Q_k(x_k)$. If the goal is to minimize the free energy with regard to $Q_k(x_k)$, the second term can be safely considered as a constant $C$. Also, when the factorization of the variational distribution is used, the third term can be rewritten as $E_{Q,(x_k)}[E_{-Q_k(x_k)}[\ln P(x, o)]]$, leading to

$$D_{KL} \left[ Q(x) \left| \left| P(o, x) \right. \right. \right] = E_{Q,(x_k)}[\ln Q_k(x_k)] - E_{Q_k(x_k)}[E_{-Q_k(x_k)}[\ln P(x, o)]] + C$$

$$= E_{Q_k(x_k)}[\ln Q_k(x_k) - E_{-Q_k(x_k)}[\ln P(x, o)]] + C$$

$$\triangleq E_{Q_k(x_k)}[\ln Q_k(x_k) - \ln Q^*_k(x_k)] + C$$

$$= D_{KL} \left[ Q_k(x_k) \left| \left| Q^*_k(x_k) \right. \right. \right] + C,$$

where $\triangleq$ means equal by definition, and $\ln Q^*_k(x_k) \triangleq E_{-Q_k(x_k)}[\ln P(x, o)]$. The KL divergence cannot be negative, which means that $Q_k(x_k) = Q^*_k(x_k)$ minimizes the free energy, and for this reason, $Q^*_k(x_k)$ is called the optimal posterior.

### 6.3 Variational Message Passing Updates

Restarting with the definition of $Q^*_k(x_k)$ and using the factorization of the generative model, we get

$$\ln Q^*_k(x_k) \triangleq E_{-Q_k(x_k)}[\ln P(x, o)]$$

$$= E_{-Q_k(x_k)}[\ln \prod_i P(N_i | p_a_i)],$$

where $N_i$ iterates over all nodes, that is, all latent and observed variables, and $p_a_i$ are the parents of $N_i$. The term in the above product can be classified into three groups: the terms that do not depend on $x_k$, the terms whose target variable ($N_i$) is $x_k$, and the terms whose predictors ($p_a_i$) contains $x_k$. Building on this observation, one can use the log property and the linearity of expectation to isolate the terms that depend on $x_k$:
Figure 12: The Markov blanket of node $A$, which is drawn in gray surrounded by a dashed line. The nodes $F$ and $G$ are the parents of $A$, and the nodes $C$ and $D$ are the children of $A$. The node $E$ is the co-parent of $A$ with respect to $D$, and the node $B$ is the co-parent of $A$ with respect to $C$.

\[
\ln Q_k^*(x_k) = \langle \ln \prod_i P(N_i|p_i) \rangle_{Q_k} \\
= \langle \ln P(x_k|p_k) \rangle_{Q_k} + \sum_{c_j \in c_k} \langle \ln P(c_j|x_k, c_{kj}) \rangle_{Q_k} + C, \quad (6.2)
\]

where $\langle \cdot \rangle_{Q_k}$ is just another notation for $E_{Q_k}(\cdot)$, and the constant $C$ comes from the terms of the product that do not depend on $x_k$. Equation 6.2 is the variational message passing equation that tells us how to compute the optimal posterior of any hidden state $x_k$ based on its Markov blanket: $x_k$’s parents $p_k$, children $c_k$, and co-parents $c_{kj}$. Figure 12 provides a visual depiction of the underlying notion of Markov blankets.

### 6.4 Conjugate Exponential Model

The variational message passing algorithm can be derived for the class of conjugate exponential models (Winn & Bishop, 2005). Those models have a likelihood function and a prior in the exponential family. Furthermore, the prior and the likelihood are conjugate, meaning that the posterior will have the same form as the prior. We follow the steps in Winn and Bishop, while referring interested readers to Winn and Bishop (2005) for more details. The derivations in equations 6.3 to 6.7 are clarified in the example in Figure 13.

Returning to our goal of computing the posterior over $x_k$ (see equation 6.2), we assume that $P(x_k|p_k)$ and $P(c_j|x_k, c_{kj})$ are in the exponential family,

\[
\ln P(x_k|p_k) = \mu_k(p_k) \cdot u_k(x_k) + h_k(x_k) + z_k(p_k), \quad (6.3)
\]
\[
\ln P(c_j|x_k, c_{kj}) = \mu_j(x_k, c_{kj}) \cdot u_j(c_j) + h_j(c_j) + z_j(x_k, c_{kj}), \quad (6.4)
\]
Realizing Active Inference in Variational Message Passing

Figure 13: This figure illustrates the computation of the optimal posterior parameters for the variable $Y$ as a message passing procedure, which requires the transmission of messages from the parent ($m_2$) and child ($m_3$) factors. Additionally, the message from the child factor ($m_3$) requires the computation of messages from the co-parent ($m_4$) and child ($m_5$) variables. Also, the message from the parent ($m_2$) factor requires the computation of a message ($m_1$) from the parent variable. Set notation and associated brackets $[]$ have been dropped, since there is only ever one parent or co-parent.

where $\mu_k(p_k), u_k(x_k), h_k(x_k)$, and $z_k(p_k)$ are the parameters, the sufficient statistics, the underlying measure, and the log partition, respectively. For a specific example, equation 7.1 shows the Dirichlet distribution written in the form of the exponential family. The first step of the Winn and Bishop method takes advantage of the conjugacy constraint to re-arrange equation 6.4 as a function of $u_k(x_k)$ that appears in equation 6.3:

$$\ln P(c_j|x_k, c_{kj}) = \mu_{j->k}(c_j, c_{kj}) \cdot u_k(x_k) + \lambda(c_j, c_{kj}),$$

where $\mu_{j->k}(c_j, c_{kj})$ and $\lambda(c_j, c_{kj})$ emerge from the rearrangement. For a specific example of this first step, see the derivation from equation 7.2 to 7.3, Figure 13 also provides an example of $\mu_{j->k}(c_j, c_{kj})$. The second step substitutes equations 6.5 and 6.3 within the variational message passing equation leading to

$$\ln Q_k^*(x_k) = (\mu_k(p_k) \cdot u_k(x_k) + h_k(x_k) + z_k(p_k)) \cdot Q_k + \sum_{c_j \in ch_k} (\mu_{j->k}(c_j, c_{kj}) \cdot u_k(x_k) + \lambda(c_j, c_{kj})) \cdot Q_k + \text{Const.}$$

The third step relies on taking the exponential of both sides, using the linearity of expectation, and factorizing by $u_k(x_k)$ to obtain
\[
Q^* \mathcal{I}(x_{\mathcal{I}}) = \exp \left\{ \left[ \langle \mu_{\mathcal{I}}(p_{\mathcal{I}}) \rangle_{\mathcal{Q}_{\mathcal{I}}} + \sum_{c_j \in ch_{\mathcal{I}}} \langle \mu_{j \rightarrow \mathcal{I}}(c_j, cp_{j\mathcal{I}}) \rangle_{\mathcal{Q}_{j\mathcal{I}}} \right] \cdot u_{\mathcal{I}}(x_{\mathcal{I}}) + h_{\mathcal{I}}(x_{\mathcal{I}}) + \text{Const} \right\},
\]

where the above constant just absorbed \( z_{\mathcal{I}}(p_{\mathcal{I}}) \) and \( \lambda(c_j, cp_{j\mathcal{I}}) \), which does not depend on \( x_{\mathcal{I}} \). At this point, we already see that the prior equation 6.3, and the approximate posterior, equation 6.6, have the same functional form, that is, only their parameters differ. The fourth step reparameterizes \( \mu_{\mathcal{I}}(p_{\mathcal{I}}) \) and \( \mu_{j \rightarrow \mathcal{I}}(c_j, cp_{j\mathcal{I}}) \) in terms of the expectation of the sufficient statistics of the children, parents, and the co-parents:

\[
Q^* \mathcal{I}(x_{\mathcal{I}}) = \exp \left\{ \mu^*_{\mathcal{I}} \cdot u_{\mathcal{I}}(x_{\mathcal{I}}) + h_{\mathcal{I}}(x_{\mathcal{I}}) + \text{Const} \right\},
\]

\[
\mu^*_{\mathcal{I}} = \tilde{\mu}_{\mathcal{I}}(\langle u_i(i) \rangle_{\mathcal{Q}_i}, \{ \langle u_l(l) \rangle_{\mathcal{Q}_l} \}_{l \in cp_{\mathcal{I}X}}),
\]

where \( \tilde{\mu}_{\mathcal{I}} \) is a reparameterization of \( \mu_{\mathcal{I}}(p_{\mathcal{I}}) \) in terms of the expectation of the sufficient statistic of the parents of \( x_{\mathcal{I}} \), and similarly, \( \tilde{\mu}_{j \rightarrow \mathcal{I}} \) is a reparameterization of \( \mu_{j \rightarrow \mathcal{I}}(c_j, cp_{j\mathcal{I}}) \). The exact form of \( \tilde{\mu}_{\mathcal{I}} \) and \( \tilde{\mu}_{j \rightarrow \mathcal{I}} \) varies from distribution to distribution. An example of those reparameterizations is visible from equations 7.4 to 7.5.

To understand the intuition behind equation 6.7, let us consider the following example: given the Forney factor graph illustrated in Figure 13, we wish to compute the posterior of \( Y \). Then the only parent of \( Y \) is \( Z \), the only child of \( Y \) is \( X \), and the only co-parent of \( Y \) with respect to \( X \) is \( W \). Therefore, applying equation 6.7 to our example leads to the equation presented in Figure 13 whose components can be interpreted as messages. Indeed, each variable (i.e., \( X, Z \) and \( W \)) sends the expectation of their sufficient statistic (i.e., a message) to the square node in the direction of \( Y \) (i.e., either \( P_X \) or \( P_Y \)). Those messages are then combined using a function (i.e., either \( \tilde{\mu}_Y \) or \( \tilde{\mu}_{X \rightarrow Y} \)) whose output (i.e., another set of messages) is summed to obtain the optimal parameters \( \mu^*_Y \). The computation of the optimal parameters, equation 6.7, can then be understood as a message passing procedure.

Returning to the Winn and Bishop (2005) method, the last step computes the (set of) expectations associated with \( \{ \langle u_i(i) \rangle_{\mathcal{Q}_i} \}_{i \in cp_{\mathcal{I}Y}}, \langle u_X(X) \rangle_{\mathcal{Q}_X} \), and \( \{ \langle u_l(l) \rangle_{\mathcal{Q}_l} \}_{l \in cp_{\mathcal{I}X}} \). Because all nodes of the model are in the exponential family, the moment generating function can be used to prove the following, 

\[
\langle u_N(N) \rangle_{\mathcal{Q}_N} = -\frac{\partial z_N(\theta_N)}{\partial \theta_N},
\]

where \( N \) is any node of the graphical model, \( \theta_N \) are the natural parameters of the distribution over \( N \), and \( z_N(\theta_N) \) is a reparameterization of the log
partition with regard to the natural parameters of the distribution over $Z$. Note that another way to compute those expectations will be presented in section 7.3.

7 The Link between Active Inference and Variational Message Passing

The previous sections have presented the theory behind active inference and variational message passing. This section focuses on the link between those two frameworks. First, we slightly modify the generative model and the variational distribution. These modifications concern a small part of the generative model and ensure conjugacy between the random variables of the model. Then we derive new update equations based on the Winn and Bishop method (Winn & Bishop, 2005). As we will see, those updates can be interpreted as a passing of messages that highlight the connection between variational message passing and belief updating in (planning as) active inference.

7.1 Generative Model Modifications. In order to perform variational message passing, we have made three modifications to the generative model described by equation 5.5. First, the prior over the precision parameter $\gamma$ is removed. Second, the softmax function forming the prior over the policies is transformed into a categorical distribution with parameters $\alpha$. This is a mild modification because the softmax function is frequently used to represent a categorical distribution—for example, neural classifiers using a softmax function as output layer or similar to the updates of $Q(s_\tau)$ and $Q(\pi)$ presented in section 5.4. Finally, we assume a Dirichlet distribution over the parameters $\alpha$. Figure 14 illustrates this new generative model, where

$$P(\pi | \alpha) = \text{Cat}(\pi; \alpha),$$
$$P(\alpha) = \text{Dir}(\alpha; \theta).$$

The conjugacy between the Dirichlet and categorical distributions enables us to derive update equations that can be interpreted as messages. Recall that the prior over policies was used to bias the policy selection toward the policies that minimize expected free energy. This can be implemented in a straightforward way while preserving conjugacy by setting the parameters of the Dirichlet as follows,

$$\theta = \overrightarrow{c} - G,$$

where $G$ is the expected free energy and $\overrightarrow{c}$ is a vector of constants whose elements satisfy the following properties:
Figure 14: The new generative model obtained after replacing the gamma distribution by a Dirichlet distribution.

1. $\forall i, j : \tau_i = \tau_j$, that is, all elements are equal,
2. $\forall j : \tau_j > \max_i G_i$, that is, all $\theta_j$ are strictly positive.

To better understand the influence of $P(\alpha)$ on the selection of policies, we imagine a Dirichlet with $K$ parameters as a distribution over a $(K-1)$-simplex. Assuming that all $\theta_i$ are greater than one, the point of this simplex with the highest probability, the mode $m_\alpha$, has the following coordinates:

$$m_\alpha = \left[ \frac{\theta_i - 1}{\sum_i \theta_i - K} \right] \ldots \left[ \frac{\theta_K - 1}{\sum_i \theta_i - K} \right].$$

Studying a few special cases of the above equation sheds some light on how policy selection is influenced by $P(\alpha)$. If the $i$th numerator of the coordinates, $\theta_i - 1$, equals one and all others equal zero, then the mode $m_\alpha$ is at the corner of the simplex corresponding to the $i$th axis. If all numerators are equal to one, then the mode is at the center of the simplex. Intuitively, this means that the bigger $\theta_i$ is relative to the other $\theta_j \forall j \neq i$, the closer $m_\alpha$ is to the $i$th corner of the simplex. Additionally, the closer $m_\alpha$ is to the $i$th corner of the simplex, the more likely the $i$th policy will be. Therefore, the bigger $\theta_i$, the more likely the $i$th policy. Finally, the only part of the numerators that is not a constant is $G_i$, and the smaller $G_i$ is, the bigger the $i$th numerator. Thus, in accord with the active inference literature, $P(\alpha)$ favors policies that minimize the expected free energy.
Another perspective on this parameterization of priors over policies is to think of $\tau$ as pseudo-counts that “promote” each policy according to how often it was previously pursued, before adding (-ve) expected free energy. If these pseudo-counts are suitably small, adding expected free energy will have a greater effect in the sense that expected free energy scores the number of times each policy would be pursued. Quantitatively, this means that a difference in the expected free energy between one policy and another can now be interpreted in terms of Dirichlet parameters or pseudo-counts.

It could be argued that the Dirichlet parameterization of the prior over policies is a more natural parameterization than the gamma distribution used to explain dopamine. Furthermore, as noted above, in most applications, gamma is set to one. More important, the precision parameter is relevant only for generative models where policies entail past transitions. In look-ahead policies or tree search implementations of planning, policies concern only future states. This means the precision of prior beliefs about policies relative to posterior beliefs (based upon the evidence a particular policy is being pursued) becomes irrelevant. In this case, the Dirichlet parameterization above may be preferred.

7.2 Variational Distribution Modifications. The variational distribution presented in section 5.2 is an example of a structured variational distribution, because factors such as $Q(S_\tau, \pi) = Q(S_\tau|\pi)Q(\pi)$ model the (posterior) dependency between $S_\tau$ and $\pi$. Performing inference with such a joint distribution falls under the category of structured variational inference (Wiegerinck, 2000; Xing, Jordan, & Russell, 2012) and is not covered in this letter. Instead, we assume a fully factorized distribution such that

$$Q(S_{0:T}, \pi, A, B, D, \gamma) = Q(\pi)Q(A)Q(B)Q(D)Q(\gamma) \prod_{\tau=0}^{T} Q(S_\tau),$$

where $Q(\pi) = \text{Cat}(\pi; \tilde{\alpha})$, $Q(S_\tau) = \text{Cat}(S_\tau; \tilde{D}_\tau)$ and all the other factors remain unchanged. This is a rather severe mean-field approximation: although it allows for straightforward application of variational message passing, removing the conditional dependencies of hidden states in the future on action means the agent cannot individuate the consequences of action. Under this functional form, the expected free energy reduces to

$$G(\pi) = \sum_{\tau=1}^{T} \mathbb{E}_{Q(S_{\tau-1:B})}[H(P(S_\tau|S_{\tau-1}, B, \pi))],$$

namely, the expected conditional entropy of the hidden states. (We also refer interested readers to appendix H for a derivation of the above equation.) Intuitively, this means that good policies select actions that lead to
Figure 15: An alternative new (expandable) generative model allowing planning under active inference. In this model, the future is now a tree-like generative model whose branches correspond to the policies considered by the agent. Each edge connecting two states in the future corresponds to an action, and the nodes in light gray represent possible expansions of the current generative model.

unambiguous hidden states. This highlights a major limitation of the mean-field approximation required by the variational message passing proposed by Winn and Bishop (2005) in the context of active inference. In other words, when removing a key structure from the variational distribution, the factor over the hidden states $Q(S_t | \pi)$ no longer depends on the policy $\pi$, and most of the terms in the expected free energy become constants with regard to $\pi$. Figure 15 illustrates an alternative generative model, implementing tree search as a form of structure learning, which is not affected by this issue because the future states in this model still depend on the action undertaken by the agent. We refer readers to our companion paper (Champion, Bowman, & Grzes, 2021) for details. A related treatment that performs exact Bayesian inference by considering a slightly different generative model can be found in Friston, Da Costa, Hafner, Hesp, and Parr (2020).
Figure 16: The differences between the framework presented in section 5 that belongs to the field of structured variational inference (Bishop & Winn, 2003) denoted by (∗), and the work presented below that belongs to the field of variational message passing (Winn & Bishop, 2005) denoted by (∗). The abbreviations BPT, E, VI, and SVMP correspond to belief propagation on tree graphical models (Kschischang et al., 2001), the elimination algorithm (Cozman, 2000), variational inference (Blei, Kucukelbir, & McAuliffe, 2017), and structured (or cluster) variational message passing (Lin, Hubacher, & Khan, 2018), respectively. Importantly, note that BPT is a specific kind of belief propagation that does not involve generalized BP (Yedidia, Freeman, & Weiss, 2000) or loopy belief propagation (Murphy, Weiss, & Jordan, 2013).

Before we turn to the derivation of the messages, we highlight the differences between active inference as presented in section 5 and the current treatment. The former is an example of structured variational inference (∗). In contrast, the work presented in this section assumes a fully factorized variational distribution and will be strictly framed as a message passing algorithm, that is, variational message passing (∗). Figure 16 illustrates those differences. Finally, we present the derivation of the messages for \( D \), \( A \), \( \pi \), and \( \alpha \), and refer readers to appendixes F and G for the derivations of the messages for \( B \) and \( S_t \), respectively.

### 7.3 Messages for \( D \)

This section applies the method of Winn and Bishop discussed in section 6.4 to compute the messages of \( D \). Let us start with the definition of the Dirichlet and categorical distributions written in the form of the exponential family:

\[
\ln P(D; d) = \begin{bmatrix}
\ln d_1 - 1 \\
\vdots \\
\ln d_{|S|} - 1
\end{bmatrix} + \begin{bmatrix}
\ln D_1 \\
\vdots \\
\ln D_{|S|}
\end{bmatrix} - \ln B(d), \quad (7.1)
\]
\[ \ln P(S_0; D) = \begin{bmatrix} \ln D_1 \\ \vdots \\ \ln D_{|S|} \end{bmatrix} \begin{bmatrix} [S_0 = 1] \\ \vdots \\ [S_0 = |S|] \end{bmatrix} \begin{bmatrix} \mu_{S_0}(D) \\ \mu_{S_0\rightarrow D}(S_0) \end{bmatrix} \cdot \begin{bmatrix} \ln D_1 \\ \vdots \\ \ln D_{|S|} \end{bmatrix} \begin{bmatrix} [S_0 = 1] \\ \vdots \\ [S_0 = |S|] \end{bmatrix}, \tag{7.2} \]

where \( B(d) \) is the beta function and \( |S| \) is the number of values a hidden state can take. The first step requires us to rewrite equation 7.2 as a function of \( u_D(D) \), this is straightforward because \( \mu_{S_0}(D) \) is just another name for \( u_D(D) \). Using the fact that the inner product is commutative,

\[ \ln P(S_0; D) = \begin{bmatrix} \ln D_1 \\ \vdots \\ \ln D_{|S|} \end{bmatrix} \begin{bmatrix} [S_0 = 1] \\ \vdots \\ [S_0 = |S|] \end{bmatrix} \begin{bmatrix} \mu_{S_0}(D) \\ \mu_{S_0\rightarrow D}(S_0) \end{bmatrix} \cdot \begin{bmatrix} \ln D_1 \\ \vdots \\ \ln D_{|S|} \end{bmatrix} \begin{bmatrix} [S_0 = 1] \\ \vdots \\ [S_0 = |S|] \end{bmatrix}, \tag{7.3} \]

The second step aims to substitute equations 7.1 and 7.3 within the variational message passing equation 6.2, that is,

\[ \ln Q^*(D) = \begin{bmatrix} \ln D_1 \\ \vdots \\ \ln D_{|S|} \end{bmatrix} \begin{bmatrix} [S_0 = 1] \\ \vdots \\ [S_0 = |S|] \end{bmatrix} + \text{Const}, \tag{7.4} \]

where \( \langle \cdot \rangle \) refers to \( \langle \cdot \rangle _{\sim Q_D} \). Note that in the above equation, \( d_i \) are fixed parameters; therefore, there is no posterior over \( d \), and the first expectation \( \langle \cdot \rangle _{\sim Q_D} \) can be removed. The third step rests on taking the exponential of both sides, using the linearity of expectation and factorizing by \( u_D(D) \) to obtain

\[ Q^*(D) = \exp \left\{ \begin{bmatrix} d_1 - 1 + \langle [S_0 = 1] \rangle \\ \vdots \\ d_{|S|} - 1 + \langle [S_0 = |S|] \rangle \end{bmatrix} \cdot u_D(D) + \text{Const} \right\}, \tag{7.4} \]

where \( z_D(d) \) have been absorbed into the constant term because it does not depend on \( D \). The fourth step is a reparameterization done by observing...
that \( [S_0 = i] \) is the \( i \)th element of the expectation of the vector \( u_{S_0}(S_0) \), that is, \( \langle u_{S_0}(S_0) \rangle_i = [S_0 = i] \):

\[
Q^*(D) = \exp \left\{ \begin{bmatrix}
    d_1 - 1 + \langle u_{S_0}(S_0) \rangle_1 \\
    \vdots \\
    d_{|S|} - 1 + \langle u_{S_0}(S_0) \rangle_{|S|}
\end{bmatrix} \cdot u_D(D) + \text{Const} \right\}. \tag{7.5}
\]

The last step consists of computing the expectation of \( \langle u_{S_0}(S_0) \rangle_i \) for all \( i \). This can be achieved by realizing that the probability of an indicator function for an event is the probability of this event:

\[
\langle u_{S_0}(S_0) \rangle_i = \langle [S_0 = i] \rangle = Q(S_0 = i) = \tilde{D}_0.\]

Substituting this result in equation 7.5, leads to the final result:

\[
Q^*(D) = \exp \left\{ \begin{bmatrix}
    d_1 - 1 + \tilde{D}_0 \\
    \vdots \\
    d_{|S|} - 1 + \tilde{D}_0
\end{bmatrix} \cdot u_D(D) + \text{Const} \right\}. \tag{7.5}
\]

Indeed, the above equation is in fact a Dirichlet distribution in exponential family form and can be rewritten into its usual form to obtain the final update equation:

\[
Q^*(D) = \text{Dir}(D; d + \tilde{D}_0).\]

In the following sections, we provide derivations for the messages of \( A, B, \pi, \alpha, \) and \( S_\tau \). Those derivations are similar to the one presented above. We encourage technical readers to go through those derivations because they constitute the main contribution of this letter. However, a reader uninterested in the algebraic details of the proofs may want to jump to section 7.7.

### 7.4 Messages for A

In the previous section, we showed how to compute the messages for \( D \), which are based on the conjugacy between a categorical \( P(S_0|D) \) and a Dirichlet \( P(D; d) \) distribution. In this section, we dive into the derivation of the messages of \( A \), which relies on the same kind of conjugacy. We start with the definition of \( P(A; a) \), which is a product of Dirichlet distributions. This product can be turned into a sum by taking the logarithm of both sides and using the log property to obtain

\[
\ln P(A; a) = \ln \prod_i P(A_i; a_i) = \sum_i \ln \text{Dir}(A_i; a_i)
\]
\[
\sum_i \left[ a_{1i} - 1 \right] \left[ \ln A_{1i} \right] - \ln B(a_i) = \sum_i \left[ a_{(O)}_{ij} - 1 \right] \left[ \ln A_{(O)}_{ij} \right] - \ln B(a_j), \tag{7.6}
\]

where \( |O| \) is the number of possible outcomes. Note that the vectors \( u_A(A) \) and \( \mu_A(a) \) step through all the elements of the matrices \( A \) and \( a \), respectively. Also, for each time step \( \tau \) up to the present time \( t \), the random matrix \( A \) has one child \( O_\tau \) (see Figure 14), and its probability mass function \( P(O_\tau | A, S_\tau) \) is a product of categorical distributions that can be written as

\[
\ln P(O_\tau = k | A, S_\tau = l) = \ln B_{[U_{l-1}]}^{|S|} - \sum_{i,j,k,u} [S_\tau = i][S_{\tau-1} = j][\pi = k][U_{l-1}^k = u] \ln B[u]_{ij}
\]

Finally, the reparameterization in the fourth step will require the probability mass function of \( S_\tau \) (see Figure 14), that is, the co-parent of \( A \) with respect to \( O_\tau \), to be written in the form of the exponential family as follows:

\[
\ln P(S_\tau = k | B, S_{\tau-1} = l, \pi = m) = \ln B[U_{l-1}^m]_{kl}
\]

\[
\sum_{i,j,k,u} [S_\tau = i][S_{\tau-1} = j][\pi = k][U_{l-1}^k = u] \ln B[u]_{ij}
\]

\[
\mu_{S_\tau}(B, S_{\tau-1}, \pi) \cdot u_{S_\tau}(S_\tau), \tag{7.8}
\]

where

\[
\mu_{S_\tau}(B, S_{\tau-1}, \pi) = \left[ \sum_{j,k,u} [S_{\tau-1} = j][\pi = k][U_{l-1}^k = u] \ln B[u]_{lj} \right], \quad \ldots, \quad \left[ \sum_{j,k,u} [S_{\tau-1} = j][\pi = k][U_{l-1}^k = u] \ln B[u]_{lj} \right]
\]
Realizing Active Inference in Variational Message Passing

and

\[ u_S(S) = \begin{bmatrix} [S = 1] \\ \vdots \\ [S = |S|] \end{bmatrix}. \]

The first step requires us to rewrite equation 7.7 as a function of \( u_A(A) \), which is done by expanding the inner product and rearranging:

\[ \ln P(O|A, S) = \begin{bmatrix} [O = 1][S = 1] \\ \vdots \\ [O = |O|][S = |S|] \end{bmatrix} \cdot u_A(A) \cdot \begin{bmatrix} \ln A_{11} \\ \vdots \\ \ln A_{|O||S|} \end{bmatrix}. \] (7.9)

The second step aims to substitute equations 7.6 and 7.9 within the variational message passing equation 6.2,

\[ \ln Q^*(A) = \begin{bmatrix} a_{11} - 1 \\ \vdots \\ a_{|O||S|} - 1 \end{bmatrix} \cdot u_A(A) \]

\[ + \sum_{\tau=0}^{t} \begin{bmatrix} [O_\tau = 1][S_\tau = 1] \\ \vdots \\ [O_\tau = |O|][S_\tau = |S|] \end{bmatrix} \cdot u_A(A) + \text{Const}, \]

where \( \langle \cdot \rangle \) refers to \( \langle \cdot \rangle_{-Q_\tau} \). The third step builds on this equation by pulling the sum over all time steps \( \tau \) inside the vector using the linearity of expectation, factorizing \( u_A(A) \), and taking the exponential of both sides:

\[ Q^*(A) = \exp \begin{bmatrix} a_{11} - 1 + \sum_{\tau=0}^{t} \langle [O_\tau = 1][S_\tau = 1] \rangle \\ \vdots \\ a_{|O||S|} - 1 + \sum_{\tau=0}^{t} \langle [O_\tau = |O|][S_\tau = |S|] \rangle \end{bmatrix} \cdot u_A(A) + \text{Const}, \]

where we used that \( a_{ij} \) are hyperparameters that are constant with regard to the expectation \( \langle \cdot \rangle_{-Q_\tau} \). The fourth step consists of two reparameterizations performed by observing that \( \langle [O_\tau = j] \rangle \) and \( \langle [S_\tau = i] \rangle \) are the expectations of the \( j \)th and \( i \)th elements of the vectors \( u_O(O) \) and \( u_S(S) \), respectively (see equations 7.7 and 7.8). Substituting those reparameterizations in the
above equation leads to

\[
Q^*(A) = \exp \left\{ a_{11} - 1 + \sum_{\tau=0}^{t} (u_{O_\tau}(O_\tau))_{11} (u_{S_\tau}(S_\tau))_{11} \right\} \cdot u_A(A) + \text{Const} \right\}.
\]

(7.10)

The last step consists of computing the expectation of \( u_{O_\tau}(O_\tau) \) and \( (u_{S_\tau}(S_\tau))_{ij} \) for all \( i \) and \( j \). Since the probability of an indicator function for an event is the probability of this event, we are searching for the probabilities of \( O_\tau = j \) and \( S_\tau = i \). The probability of \( O_\tau = j \) is the \( j \)th element of the vector \( a_\tau \), which is a one hot vector containing the observation from the environment at time \( \tau \). The posterior probability of \( S_\tau = i \) is by definition \( Q(S_\tau) = D_\tau \). Substituting the probabilities of \( O_\tau = j \) and \( S_\tau = i \) in equation 7.10 leads to

\[
Q^*(A) = \exp \left\{ a_{11} - 1 + \sum_{\tau=0}^{t} a_{\tau j} D_{\tau 1} \right\} \cdot u_A(A) + \text{Const} \right\}.
\]

(7.11)

\[
= \prod_i \exp \left\{ a_{ij} - 1 + \sum_{\tau=0}^{t} a_{\tau i} D_{\tau j} \right\} \cdot \ln A_{ij} + \text{Const} \right\}.
\]

(7.12)

Finally, one can recognize in equation 7.12 the product of Dirichlet distributions written into their exponential form:

\[
Q^*(A) = \prod_i \text{Dir}(A_{\cdot i}, a_{\cdot i}) \text{ where } a = a + \sum \alpha_\tau \otimes D_\tau.
\]

The origin of the outer product in the computation of the parameters can be understood by considering \( P' \) the outer product between \( \alpha_\tau \) and \( s_\tau \) such that \( P'_{ij} = \alpha_\tau \otimes s_\tau \). Then, equation 7.11 shows that \( a_{ij} = a_{ij} + \sum P'_{ij} \Leftrightarrow a = a + \sum \alpha_\tau \otimes s_\tau \).

### 7.5 Messages for \( \pi \)

We now turn to the messages for \( \pi \). Note, that the definitions of the \( P(S_\tau | B, S_{\tau-1}, \pi) \) and \( P(\pi | a) \) are given by equations 7.8 and F.3, respectively. The first step requires us to rewrite equation 7.8 as a function of \( u_\tau(\pi) \). Using the inner product definition and rearranging, we
Realizing Active Inference in Variational Message Passing

obtain

\[
\ln P(S_t = k|B, S_{t-1} = l, \pi = m) = \left[ \sum_{i,j,u} [S_t = i][U_{t-1}^1 = u][S_{t-1} = j] \ln B[u_{ij}] \right] u_{\pi}(\pi). \tag{7.13}
\]

The second step aims to substitute equations 7.3 and 7.13 within the variational message passing equation,

\[
\ln Q^*(\pi) = \left[ \ln \alpha_1 \right] \cdot u_\pi(\pi) + \sum_{t=1}^T \left[ \sum_{i,j,u} [S_t = i][U_{t-1}^1 = u][S_{t-1} = j] \ln B[u_{ij}] \right] u_\pi(\pi) + \text{Const},
\]

where \( \cdot \) refers to \( \langle \cdot \rangle_{Q^*} \). The third step relies on pulling the summation over all time steps inside the vector, taking the exponential of both sides, using the linearity of expectation, and factorizing by \( u_\pi(\pi) \) to obtain

\[
Q^*(\pi) \propto \exp \left\{ \left[ \ln \alpha_1 \right] + \sum_{t,i,j,u} [U_{t-1}^1 = u] \langle [S_t = i][S_{t-1} = j] \rangle \ln B[u_{ij}] \right\} u_\pi(\pi).
\]

The fourth step is a reparameterization implemented by observing that \( \langle \ln \alpha_1 \rangle, \langle [S_t = i] \rangle, \langle [S_{t-1} = j] \rangle \), and \( \langle \ln B[u_{ij}] \rangle \) are elements of the vectors \( \langle u_\alpha(\alpha) \rangle, \langle u_{S_t}(S_t) \rangle, \langle u_{S_{t-1}}(S_{t-1}) \rangle \), and \( \langle u_B(B) \rangle \), respectively:

\[
\mu_\pi^* = \begin{cases} 
\left[ \langle u_\alpha(\alpha) \rangle_1 + \sum_{t,i,j,u} [U_{t-1}^1 = u] \langle u_{S_t}(S_t) \rangle \langle u_{S_{t-1}}(S_{t-1}) \rangle \langle u_B(B) \rangle \right] u_{i,j} \\
\left[ \langle u_\alpha(\alpha) \rangle_{\pi} + \sum_{t,i,j,u} [U_{t-1}^1 = u] \langle u_{S_t}(S_t) \rangle \langle u_{S_{t-1}}(S_{t-1}) \rangle \langle u_B(B) \rangle \right] u_{i,j} 
\end{cases}
\tag{7.14}
\]

The last step consists of computing the expectation of \( \langle u_\alpha(\alpha) \rangle_k, \langle u_{S_t}(S_t) \rangle, \langle u_{S_{t-1}}(S_{t-1}) \rangle_j \), and \( \langle u_B(B) \rangle_{u,i,j} \) for all \( i, j, k \), and \( u \):
\[ \begin{align*}
\langle u_\alpha \rangle_k &= \langle \ln \alpha_k \rangle = \psi(\tilde{\alpha}_k) - \psi(\sum_i \tilde{\alpha}_i) = \bar{\alpha}_k \\
\langle u_{S_i} \rangle_i &= \langle [S_i = i] \rangle = D_{ti} \\
\langle u_{S_{\tau-1,i}} \rangle_j &= \langle [S_{\tau-1} = j] \rangle = D_{(\tau-1)j} \\
\langle u_{B_{i,j}} \rangle_{i,j} &= \langle \ln B_{i,j} \rangle = \psi(b_{i,j}) - \psi(\sum_l b_{l,j}) = \bar{B}_{i,j}
\end{align*} \]

Furthermore, the indicator function in the \( k \)th row of equation \( 7.14 \) filters out all elements where \( u \neq U_{k-1} \). Substituting those results in equation \( 7.14 \) leads to the final result:

\[ Q^*(\pi) \propto \exp \begin{bmatrix}
\bar{\alpha}_1 + \sum_{i,j} \bar{D}_{\tau,\tau-1} B_{i,j} \bar{U}_{\tau-1,i}^1 \\
\bar{\alpha}_{|\pi|} + \sum_{i,j} \bar{D}_{\tau,\tau-1} B_{|\pi|,\tau-1} \bar{U}_{\tau-1,i}^{|\pi|}\end{bmatrix} \cdot u_\pi(\pi)
\]

Indeed, the above equation is a categorical distribution in the exponential family form and can be rewritten into its usual form as follows:

\[ Q^*(\pi) = \text{Cat}(\pi; \alpha^*) \quad \text{where} \quad \alpha^* = \bar{\alpha} + \sum_{\tau=1}^T \mathbb{F}_\tau \quad \text{and} \]

\[ \mathbb{F}_\tau = \begin{bmatrix}
\langle D_{\tau,\tau-1} B_{i,j} \bar{U}_{\tau-1,i}^1 \rangle_F \\
\langle \bar{D}_{\tau,\tau-1} B_{|\pi|,\tau-1} \bar{U}_{\tau-1,i}^{|\pi|} \rangle_F
\end{bmatrix},
\]

where it should be stressed that \( \langle \cdot, \cdot \rangle_F \) is not an expectation but the Frobenius product, that is, a generalization of the inner product to matrices.

### 7.6 Messages for \( \alpha \)

In this section, we focus on the messages for \( \alpha \), whose derivation is identical to the messages of \( D \). To see this, note that \( P(D) \) was a Dirichlet with parameters \( d \). Furthermore, the only child of \( D \) was \( S_0 \), whose prior and posterior were categorical distributions with parameters \( D \) and \( \tilde{D} \). Similarly, note that \( P(\alpha) \) is a Dirichlet with parameters \( \theta \). Furthermore, the only child of \( \alpha \) is \( \pi \) whose prior and posterior are categorical distributions with parameters \( \alpha \) and \( \bar{\alpha} \). From this observation, we directly obtain the following result:

\[ Q^*(\alpha) = \text{Dir}(\alpha; \theta + \bar{\alpha}). \]

### 7.7 Summary of Messages

Next, we focus on explaining the intuition behind the resulting equations, which can be understood as a summation of two major kinds of messages. First are the messages from the parent factors that correspond to messages of type \( m_2 \) in Figure 13. Those messages are functions of the expectation of the sufficient statistic of the parent variables, that is, functions of messages of type \( m_1 \). Second are the messages from the
Realizing Active Inference in Variational Message Passing

child factors, which correspond to messages of type $m_3$ in Figure 13. Those messages are functions of the sufficient statistics of the co-parent and child variables, that is, functions of messages of type $m_4$ and $m_5$, respectively. Let’s see how these play out in our newly derived equations.

**Messages for $\alpha$:**

$$Q^*(\alpha) = \text{Dir}(\alpha; \theta + \tilde{\alpha}).$$

Recall that $\mu_\alpha = \theta$ is an $m_2$ message (orange colour). However, $\alpha$ does not have any parent variables; thus, $\mu_\alpha$ is a constant, that is, a function of zero $m_1$ messages. Furthermore, we know that $\alpha$ has only one child variable ($\pi$) and no co-parent variables. Therefore, $\mu_{\pi \to \alpha}(\tilde{\alpha}) = \tilde{\alpha}$ is the only $m_3$ message (purple colour) for $\alpha$, where $\tilde{\alpha} = \langle u_{\pi}(\pi) \rangle_{\pi}$ is an $m_5$ message.

**Messages for $D$:**

$$Q^*(D) = \text{Dir}(D; d + \tilde{D}_0).$$

Similarly for the messages of $\alpha$, $\mu_D = d$ and $\mu_{S_\tau \to D}(\tilde{D}_0) = \tilde{D}_0$, where $\tilde{D}_0$ should be thought of as a message from a child variable ($m_5$ message).

**Messages for $A$:**

$$Q^*(A) = \prod_i \text{Dir}(A_{i}, a_{i}) \text{ where } a = a + \sum_{\tau} a_{t} \otimes D_{\tau}.$$

Following the same reasoning, $\mu_A = a$ is an $m_2$ message, and because $A$ does not have any parent variables, $\mu_A$ is a constant. Also, $A$ has one child variable ($O_\tau$) for each time step $\tau \in \llbracket 0, t \rrbracket$ and one co-parent variable ($S_\tau$) for each of them, which implies that there are $t + 1$ $m_3$ messages for $A$, that is, $\mu_{O_\tau \to A}(a_{t}, D_{\tau}) = a_{t} \otimes D_{\tau}$ for all $\tau \in \llbracket 0, t \rrbracket$. Because the $O_\tau$ are observed, we know that the $m_3$ messages transmitted by this node will be the observation made at time $\tau (a_{t})$. Additionally, the $m_4$ message from the hidden variables $S_\tau$ are the expectation of their sufficient statistics, that is, $\langle u_{S_\tau}(S_\tau) \rangle_{O_\tau} = D_{\tau}$. This confirms the idea that $\mu_{O_\tau \to A}$ is a function of the sufficient statistics of the child and co-parent variables. Figure 17 concludes this paragraph with a visual representation of the messages for $A$.

**Messages for $B$:**

$$Q^*(B) = \prod_{u,i} \text{Dir}(B[u], b[u]) \text{ where } b[u] = b[u] + \sum_{(k,r) \in \Omega_u} \hat{a}_{k} \tilde{D}_{r} \otimes \tilde{D}_{r-1}.$$
Figure 17: The passing of messages required to update the posterior over $A$. The messages of type $m_2, m_3, m_4,$ and $m_5$ come from the parent factors, child factors, co-parent variables, and child variables, respectively.

Sticking with this reasoning, $\mu_B = b$ is an $m_2$ message, and because $B$ does not have any parent variables, then $\mu_B$ is a constant equal to $b$. Also, $B$ has one child variable ($S_\tau$) for each time step $\tau \in [1, T]$ and all policies $\forall \pi \in [1, |\pi|]$, along with two co-parent variables ($S_{\tau-1}$ and $\pi$) for each of those child variables. This implies that there are $T \times |\pi| m_3$ messages for $B$, that is, $\mu_{S_\tau \rightarrow S_{\tau-1}}(\tilde{\alpha}_k, \tilde{D}_\tau \otimes \tilde{D}_{\tau-1}) = \hat{\alpha}_k \tilde{D}_\tau \otimes \tilde{D}_{\tau-1}$. $\forall \tau \in [1, T] \ \forall \pi \in [1, |\pi|]$ where $\tilde{D}_\tau$ is an $m_5$ message and $\tilde{\alpha}_k$ along with $\tilde{D}_{\tau-1}$ are $m_4$ messages.

**Messages for $\pi$:**

$$Q^*(\pi) = \text{Cat}(\pi; \alpha^*) \ \text{where} \ \alpha^* = \sigma \left( \bar{a} + \sum_{\tau=1}^T \mathbb{F}_\tau \right)$$

$$\mathbb{F}_\tau = \begin{bmatrix} \langle \tilde{D}_\tau \otimes \tilde{D}_{\tau-1}, B[U^1_{\tau-1}] \rangle_f \\ \vdots \\ \langle \tilde{D}_\tau \otimes \tilde{D}_{\tau-1}, B[U^{|\pi|}_{\tau-1}] \rangle_f \end{bmatrix}.$$  

If we keep applying the same reasoning, we see that $\mu_\pi(\bar{a}) = \bar{\alpha}$ is an $m_2$ message, which is a function of the sufficient statistics of the parent variable $\alpha$ ($m_1$ message). Moreover, $\pi$ has one child variable ($S_\tau$) for each time step $\tau \in [1, T]$, and for each of those child variables, $\pi$ has two co-parent variables ($S_{\tau-1}$ and $B$). Therefore, $\mu_{S_{\tau} \rightarrow \pi} = \mathbb{F}_\tau \ \forall \tau \in [1, T]$ correspond to $T m_3$.
Realizing Active Inference in Variational Message Passing

messages. Those messages are function of two $m_4$ messages ($\tilde{D}_{\tau - 1}$ and $\bar{B}$) and one $m_5$ message ($D_{\tau}$).

**Messages for $S_\tau$:**

$$Q^*(S_{\tau}) = \text{Cat}(S_{\tau}; \sigma(\mu_{S_{\tau}}))$$

$$\mu_{S_{\tau}} = [\tau = 0 | D + [\tau \neq 0] \sum_k \tilde{\alpha}_k B[U_k^{\tau}] | D_{\tau - 1} + [\tau \leq t] o_{\tau} \cdot \bar{A}$$

$$+ [\tau \neq T] \sum_k \tilde{\alpha}_k D_{\tau + 1} \cdot B[U_k^{\tau}]$$

To understand the above equation, we can consider two cases: $\tau = 0$ and $\tau \neq 0$. In the first case, $S_0$ has only one parent variable ($D$), and $\mu_{S_0}(D) = D$ where $D = (\mu_D(D))_{|Q_0}$ is a message from a parent variable ($m_1$ message).

In the second case, $S_\tau$ has three parent variables ($S_{\tau - 1}, B$ and $\pi$), and $\mu_{S_\tau}(D_{\tau - 1}, B, \bar{\alpha}) = \sum_k \tilde{\alpha}_k B[U_k^{\tau}] D_{\tau - 1}$ where $D_{\tau - 1}$, $B$, and $\bar{\alpha}$ are also $m_1$ messages. Let us now think about the child variable(s) of $S_\tau$. If $\tau \leq t$, then $S_\tau$ has a child variable from the likelihood mapping and $\mu_{O_{\tau} \rightarrow S_\tau}(o_{\tau}, \bar{A}) = o_{\tau} \cdot \bar{A}$, where $o_{\tau}$ is a message from the child variable ($m_5$ message) and $\bar{A}$ is a message from the co-parent variable ($m_4$ message). Additionally, if $\tau \neq T$, then $S_\tau$ receives a message from the future $\mu_{S_{\tau + 1} \rightarrow S_\tau}(\tilde{\alpha}_k, D_{\tau + 1}, B) = \sum_k \tilde{\alpha}_k D_{\tau + 1} \cdot B[U_k^{\tau}]$, where $\tilde{\alpha}_k$ and $B$ are $m_4$ messages and $D_{\tau + 1}$ is a $m_3$ message. Figure 18 concludes this section with an illustration the message passing procedure for $S_0$.

**7.8 Messages versus Update Equations.** In this section, we present a side-by-side comparison of the messages obtained using variational message passing and the update equations that underwrite belief updating in the active inference literature. Throughout this section, the messages always represented first, followed by the equivalent update equations. We start with the random variable $D$:

$$Q^*(D) = \text{Dir}(D; d + \tilde{D}_0),$$

$$Q^*(D) = \text{Dir}(D; d + s_0).$$

These two equations differ only in terms of labels: $s_0$ and $D_0$ conceptually represent the same quantity. Similarly, the updates of $A$ are recovered up to a change of label:

$$Q^*(A) = \prod_i \text{Dir}(A_{s_i}, a_i) \quad \text{where} \quad a = a + \sum_{\tau=0}^{t} o_{\tau} \otimes D_{\tau},$$

$$Q^*(A) = \prod_i \text{Dir}(A_{s_i}, a_i) \quad \text{where} \quad a = a + \sum_{\tau=0}^{t} o_{\tau} \otimes s_{\tau}.$$
Figure 18: The passing of messages required to update the posterior over $S_0$. The messages of types $m_1, m_2, m_3, m_4,$ and $m_5$ come from the parent variables, parent factors, child factors, co-parent variables, and child variables, respectively.

The update of $B$ slightly differs from the messages obtained from variational message passing, which follows from the fact that we modified the variational distribution:

$$Q^*(B|\sigma) = \prod_{u,i} \text{Dir}(B[u] \cdot i, b[u] \cdot i)$$

where $b[u] = b[u] + \sum_{(k, \tau) \in \Omega_\tau} \tilde{\alpha}_k \tilde{D}_\tau \otimes \tilde{D}_{\tau-1}$.

The only conceptual difference here is that $s^\tau$ depended on the policy, while $D$ does not. Concerning $S_\tau$, we have rearranged the update equation to highlight the similarity with the messages:

$$Q^*(S_\tau|\sigma) = \text{Cat}(S_\tau; \sigma(\mu^*_S)),$$

$$\mu^*_S = [\tau = 0]D + [\tau \neq 0] \sum_k \tilde{\alpha}_k \tilde{D}_{\tau-1} \cdot \tilde{B}[U_{\tau-1}^k] + [\tau \leq t] o_\tau \cdot A$$

$$+ [\tau \neq T] \sum_k \tilde{\alpha}_k \tilde{D}_{\tau+1} \cdot \tilde{B}[U_{\tau+1}^k].$$

$$\mu^*_S = [\tau = 0]D + [\tau \neq 0] \tilde{B}[U_{\tau-1}^\tau] s^\tau_{\tau-1} + [\tau \leq t] o_\tau \cdot A + [\tau \neq T] \tilde{B}[U_{\tau+1}^\tau] s^\tau_{\tau+1}.$$

$$Q^*(S_0|\sigma) = \sigma(\tilde{D}_0 \cdot A + \sum_k \tilde{\alpha}_k \tilde{D}_\tau \cdot \tilde{B}[U_{\tau}^0]).$$
Realizing Active Inference in Variational Message Passing

There are two main differences here. First, as for $B$, $s^*_t$ is replaced by $\tilde{D}$, which does not depend on the policies. Second, the past and future messages have an average over the policies, while the updates do not. Unsurprisingly, since we replaced $\gamma$ by $\alpha$ and changed the type of distributions, the updates are quite different:

\[
Q^*(\alpha) = \text{Dir}(\alpha; \theta + \bar{\alpha}), \\
Q^*(\gamma) = \Gamma(\gamma; 1, \beta + G \cdot (\pi - \pi_0)).
\]

We conclude this section with the messages and updates of $\pi$, which are formally distinct. These differences come from the fact that we moved $G$ from $P(\pi | \gamma)$ to $P(\alpha)$ and turned $P(\pi | \gamma)$ into a categorical distribution $P(\pi | \alpha)$:

\[
Q^*(\pi) = \text{Cat}(\pi; \alpha^*), \\
\alpha^* = \sigma\left( \bar{a} + \sum_{\tau=1}^{T} F_{\tau} \right) \text{ and } F_{\tau} = \begin{bmatrix} \langle \tilde{D}_{\tau} \otimes \tilde{D}_{\tau-1}, B[U^1_{\tau-1}] \rangle_F \\ \vdots \\ \langle \tilde{D}_{\tau} \otimes \tilde{D}_{\tau-1}, B[U^{\pi \tau}] \rangle_F \\ \end{bmatrix},
\]

\[
\alpha^* = \sigma\left( -\frac{1}{\beta} G + \sum_{\tau=1}^{T} F_{\tau} \right) \text{ and } F_{\tau} = s^*_\tau \cdot \tilde{B}[U] s^*_{\tau-1}. 
\]

However, the general form of the updates remains unchanged with information coming from the parent through $\bar{a}$ and $-\frac{1}{\beta} G$ and from each child through the summation over time steps.

8 Conclusion

The increasing use of active inference in neuroscience has cast many brain processes as Bayesian inference, the update equations of which can be thought of as a message passing procedure. The first goal of this letter was to present a complete overview of the active inference framework in discrete time and state space (section 5) as well as a formal introduction to the variational message passing literature (section 6). Then we simplified the generative model and the variational distribution usually adopted in the active inference to derive a new set of update equations using the method of Winn and Bishop (2005) and highlight the connection between active inference and variational message passing (section 7).

We hope that the first few sections of this letter will be useful as an introduction to variational inference, Forney factor graphs, active inference, and variational message passing. Section 7 might also be of interest to researchers searching for a clear link between active inference and variational message passing or researchers seeking to derive the update equations of new generative models. Section 7 explains why a fully factorized variational distribution simplifies the expected free energy in a way that precludes
risk-sensitive behavior but preserves ambiguity avoidance. Finally, we note that this issue does not confound generative models implementing tree search.

One might ask why previous formulations of belief updating or message passing in active inference have not exploited the simplifications considered in this letter—for example, using a Dirichlet distribution to parameterize Bayesian beliefs over policies or a fully factorized variational distribution that would simplify message passing. One answer is that much of the legacy literature in active inference is concerned with neuronal process theories and biological implementation. For example, the only reason a Gibbs form was used for the distribution over policies was to link the implicit temperature or sensitivity parameter to dopaminergic discharges. Similarly, the minimization of variational free energy—using a gradient descent to implement structured variational message passing—was motivated by the need to cast belief updating in terms of differential equations that could be plausibly associated with neuronal dynamics (and accompanying electrophysiological responses to observations). However, if one frees oneself from the constraints of biological implementation, the repertoire of established schemes in machine learning and Bayesian statistics can, in principle, be leveraged to reproduce kinds of choice behavior active inference is trying to explain and emulate. This letter has highlighted the putative usefulness of variational message passing under a rationalization of generative models.

It is interesting to consider whether the simplified expected free energy resulting from our message passing formulation of active inference can be linked in any sense to human behavior, whether normative or pathological. In particular, the free energy we have obtained reflects a very specific functional impoverishment. The full factorization that is necessary for vanilla message passing precludes the ability to conditionalize the variational posterior on policies. This suggests a particular deficit in the ability to plan and a blindness to future possibilities, the uncertainty associated with those possibilities, and their potential to satisfy preferences. As a result, the agent’s objective becomes to seek out unambiguous cues, with no concern for outcome.

In fact, humans do exhibit patterns of behavior that—due to their repetitiveness—seem to reflect a desire for high predictability. Additionally, some of these patterns do not seem obviously connected to rewarding or punishing outcomes. For example, those with autism can exhibit very stereotyped repetitive behavior: for example, hand flapping, hand clapping, or rocking (Gabriels et al., 2005), which is often described as stimming (Sundar Rajagopalan, Dhall, & Goecke, 2013). These repetitive and ritualistic behaviors (Lam, 2007) suggest an objective to avoid exploration and the associated uncertainty.

This work naturally leads to future directions of research. For example, one could implement the new generative model proposed in this letter and...
compare its performance with the model presented in section 5. Furthermore, additional research needs to be done to connect the original update equations of active inference to the cluster variational message passing literature. Much work has already been done on structured variational message passing, particularly relation to marginal message passing and its advantages over related approaches based on Bethe free energy (Yedidia, 2005; Parr et al., 2019). Another interesting direction of research would be to design new generative models that can tackle more complex tasks, such as playing Atari games, human-machine interaction using natural language, and automatic structure learning. Partial answers to these directions of research have already been provided with the use of deep active inference (Fountas et al., 2020; Ueltzhöffer, 2018; Tschantz, Baltieri, Seth, & Buckley, 2020), deep temporal models (Friston, Rosch, Parr, Price, & Bowman, 2018; Heins et al., 2020), and Bayesian model reduction (Friston, Parr, & Zeidman, 2018; Friston, Lin et al., 2017; Wauthier, Çatal, De Boom, Verbelen, & Dhoedt, 2020). Nevertheless, we anticipate that additional work will pursue these avenues of research. Finally, one could also compare the update schemes under VMP to belief propagation (Yedidia, 2011) or marginal message passing (Parr et al., 2019).

Appendix A: Active Inference, KL Control, and Reinforcement Learning

This appendix focuses on the relationship between active inference, KL control, and reinforcement learning (see DaCosta, Sajid, Parr, Friston, & Smith, 2020, and Levine, 2018, for more details). Let us restart with the expected free energy given by equation 5.4:

\[
G(\pi) \approx \sum_{t=\tau+1}^{T} D_{KL}[\tilde{Q}(O_t|\pi) \parallel \tilde{P}(O_t)] + E_{Q(S_t|\pi)}[H[P(O_t|S_t)]].
\]

If the expected ambiguity is equal to zero, then the expected free energy reduces to the expected risk, which is the cost function minimized in the KL control literature. This highlights that active inference generalizes KL control (Rawlik et al., 2013) by taking into account the ambiguity of the mapping between the hidden states and the observations. Active inference therefore selects policies leading to unambiguous states. Furthermore, the expected risk can be rewritten as
expected risk = \( D_{KL}[Q(O_t|\pi)||P(O_t)] \)
\[
= \underbrace{E_{Q(O_t|\pi)}[\ln Q(O_t|\pi)]}_{\text{negative entropy}} - \underbrace{E_{Q(O_t|\pi)}[P(O_t)]}_{\text{expected rewards}}
\]

If the negative entropy is zero, then the expected free energy reduces to the negative expected prior preference. Those preferences encode the notion of good outcomes or, equivalently, the notion of rewarding observations. This highlights why active inference can be thought of as a generalization of reinforcement learning (Mnih et al., 2013). Another view on the expected free energy is

\[
G(\tau, \pi) = (-\text{ev}) \text{ epistemic value} - \text{extrinsic value}
\]
\[
\equiv -E_{Q} [\ln Q(S_t|\pi) - \ln P(S_t|O_t, \pi)] - E_{Q} [\ln P(O_t|\pi)]. \quad (A.1)
\]

where \( \hat{Q} = P(O_t|S_t)Q(S_t) \). The extrinsic value is another term for expected prior preferences, which is equivalent to expected rewards in reinforcement learning. It is worth looking in more detail at the negative epistemic value (-EV), which differentiates the learning objectives of reinforcement learning and active inference:

\[
-EV = -E_{\hat{Q}} [\ln Q(S_t|\pi) - \ln P(S_t|O_t, \pi)]
\]
\[
\Leftrightarrow EV = E_{\hat{Q}} [\ln P(S_t|O_t, \pi) - \ln Q(S_t|\pi)].
\]

Thus, the epistemic value is approximately equal to the mutual information between \( S_t \) and \( O_t \). The mutual information encodes the expected information gain over one variable by knowing the value of another. Therefore, the epistemic value tells us how knowing future observations reduces our uncertainty over future hidden states. The following should help to see that the epistemic value is approximately equal to the mutual information between \( S_t \) and \( O_t \):

\[
I(S; O) = D_{KL} [P(S_t, O_t)||P(S_t)P(O_t)]
\]
\[
= E_{P(S_t, O_t)} [\ln P(S_t|O_t) + \ln P(O_t) - \ln P(S_t) - \ln P(O_t)]
\]
\[
= E_{P(S_t, O_t)} [\ln P(S_t|O_t) - \ln P(S_t)].
\]

Intuitively, the more an observation tells us about future states, the more valuable this observation is. The negative epistemic value from equation A.1 directly reflects this intuition and favors the policies with high mutual information. More important, equation A.1 allows the agent to compare the
information gain and the reward on the same scale, using nats from information theory. This creates a sense in which an active inference agent deals optimally with the trade-off between exploration and exploitation.

Appendix B: Useful Properties

This appendix quickly reviews the properties used throughout this letter.

Product rule: $P(X, Y) = P(X|Y)P(Y)$, where $X$ and $Y$ are random variables.

Linearity of expectation: $E_{P(Y)}[aY + b] = aE_{P(Y)}[Y] + b$, where $a$ and $b$ are constants, and $Y$ is a random variable.

Expectation of a constant: $E_{P(Y)}[a] = a$, where $a$ is a constant, and $Y$ is a random variable.

Log property: $\ln(ab) = \ln(a) + \ln(b)$, where $a$ and $b$ are real numbers.

Exponential product property: $\exp(a + b) = \exp(a)\exp(b)$, where $a$ and $b$ are real numbers.

Exponential power property: $\exp(ab) = \exp(a)^b$, where $a$ and $b$ are real numbers.

Appendix C: Definition and Justification of the Expected Free Energy

In this appendix, we focus on the definition of the expected free energy and the justification of equation 5.4. Another good resource on the subject is the “expected free energy” appendix of Smith, Friston, and Whyte (2021). For simplicity, we assume the following generative model and variational distribution:

$$P(O_{0:T}, S_{0:T}, B|\pi) = P(B)P(S_0) \prod_{\tau=1}^{T} P(S_\tau|S_{\tau-1}, B, \pi) \prod_{\tau=0}^{T} P(O_\tau|S_\tau),$$

$$Q(S_{0:T}, B|\pi) = Q(B) \prod_{\tau=0}^{T} Q(S_\tau|\pi).$$

Furthermore, we let $X = \{B, S_{0:T}\}$ denote the set of hidden variables of the model. Note that in this appendix, we restrict ourself to the hidden variables $X$, but new variables such as $A$ and $D$ can be added without changing the idea of the following derivation. Initially, the expected free energy was defined as the variational free energy conditioned on the policy:

$$G(\pi) = D_{KL} \left[ Q(X|\pi) \right| P(O_{0:T}, X|\pi) \right].$$
However, the above definition does not take into account that observations will be made in the future. To make up for this, the expected free energy can be extended as follows:

\[
G(\pi) = \mathbb{E}_Q[D_{KL} \left( Q(X|\pi) || P(O_{0:T}, X|\pi) \right)]
\]

where \( \hat{Q} \triangleq \hat{Q}(O_{t+1:T}|\pi) \). (C.1)

Since the future observations \((O_{t+1:T})\) have not been made yet, we need to predict what they could look like. This prediction relies on a predictive distribution \(\hat{Q}(O_{t+1:T}|\pi)\) that encodes our best guess about future outcomes, and is generally defined as follows:

\[
\hat{Q}(O_{t+1:T}|\pi) \triangleq \prod_{t=1}^{T} \hat{Q}(O_t|\pi),
\]

where

\[
\hat{Q}(O_t|\pi) \triangleq \sum_{S_t} \hat{Q}(O_t, S_t|\pi) \quad \text{and} \quad \hat{Q}(O_t, S_t|\pi) \triangleq P(O_t|S_t)Q(S_t|\pi).
\]

Note that the definition of \(\hat{Q}(O_{t+1:T}|\pi)\) assumes independence between time steps and \(\hat{Q}(O_t|\pi)\) is obtained by marginalization of \(\hat{Q}(O_t, S_t|\pi)\). By recalling the definition of the generative model as well as the definition of the variational distribution, we obtain the following from equation C.1:

\[
G(\pi) = \mathbb{E}_Q[D_{KL} \left( Q(O_{0:T}, B|\pi) || P(O_{0:T}, B|\pi) \right)]
\]

\[
= D_{KL} \left( Q(B) || P(B) \right) + D_{KL} \left[ Q(S_0|\pi) || P(S_0) \right]
\]

\[
+ \sum_{t=1}^{T} \mathbb{E}_{Q(S_{t-1}, B|\pi)} \left[ D_{KL} \left( Q(S_t|\pi) || P(S_t|S_{t-1}, B, \pi) \right) \right]
\]

\[
+ \sum_{t=0}^{T} \mathbb{E}_{Q(S_t|\pi)} \left[ \mathbb{H}(P(O_t|S_t)) \right]
\]

\[
+ \sum_{t=1}^{T} \mathbb{E}_{Q(S_{t-1}, B|\pi)} \left[ D_{KL} \left( Q(S_t|\pi) || P(S_t|S_{t-1}, B, \pi) \right) \right]
\]

\[
+ \mathbb{E}_{Q(S_t|\pi)} \left[ \mathbb{H}(P(O_t|S_t)) \right] .
\]

It must now be mentioned that the policy does not have much of an impact on the past and current hidden states \((S_{0:t})\). The terms relying on those states are then removed from the expected free energy to avoid unnecessary
computational costs. Additionally, the divergence between \( Q(B) \) and \( P(B) \) does not depend on the policy and can be safely ignored, leading to

\[
G(\pi) = \sum_{r=\tau+1}^T G(\pi, r),
\]

where

\[
G(\pi, r) \triangleq \mathbb{E}_{Q(S_{r-1}, B|\pi)} \left[ D_{KL} \left( Q(S_r|\pi) \mid \mid P(S_r|S_{r-1}, B, \pi) \right) \right] + \mathbb{E}_{Q(S_r|\pi)} \left[ H(P(O_r|S_r)) \right].
\]

We now focus on \( G(\pi, r) \) to bridge the gap between equations 5.4 and C.2. First, we merge the two terms of the above equation together:

\[
G(\pi, r) \triangleq \mathbb{E}_{P(O_r|S_r)Q(S_r, S_{r-1}, B|\pi)} \left[ \ln Q(S_r|\pi) - \ln P(O_r, S_r|S_{r-1}, B, \pi) \right].
\]

Then we break the second term within the expectation using the product rule. Additionally, we realize that the following equation can be obtained from the product rule:

\[
P(O_r|S_{r-1}, B, \pi) = \frac{P(O_r, S_{r-1}, B, \pi)}{P(S_{r-1}, B, \pi)} = \frac{P(S_{r-1}, B, \pi|O_r)}{P(S_{r-1}, B, \pi)} P(O_r) \approx P(O_r),
\]

where we assumed that the fraction is equal to one. Doing this assumption means that the observation \( O_r \) brings us very little information, that is, the posterior is close to the prior. Using the above result, we get

\[
G(\pi, r) \approx \mathbb{E} \left[ \ln Q(S_r|\pi) - \ln P(S_r|O_r, S_{r-1}, B, \pi) - \ln P(O_r) \right]
\]

\[
\approx \mathbb{E} \left[ \ln Q(S_r|\pi) - \ln P(S_r|O_r, S_{r-1}, B, \pi) - \ln P(O_r) \right],
\]

where the expectation is still over \( P(O_r)Q(S_r, S_{r-1}, B|\pi) \). Then we use Bayes’ theorem on the second term—the fact that \( O_r \perp S_{r-1}, B, \pi \), and the log properties to get

\[
G(\pi, r) = \mathbb{E} \left[ \ln Q(S_r|\pi) - \ln P(S_r|O_r, S_{r-1}, B, \pi) - \ln P(O_r) \right]
\]

\[
= \mathbb{E} \left[ \ln Q(S_r|\pi) - \ln \frac{P(O_r|S_r, S_{r-1}, B, \pi)P(S_r|S_{r-1}, B, \pi)}{P(O_r|S_{r-1}, B, \pi)} - \ln P(O_r) \right]
\]

\[
\approx \mathbb{E} \left[ \ln Q(S_r|\pi) - \ln \frac{P(O_r|S_r)Q(S_r|\pi)}{Q(O_r|\pi)} - \ln P(O_r) \right]
\]
\[ = \mathbb{E}\left[ \ln Q(O_t|\pi) - \ln P(O_t) - \ln P(O_t|S_t) \right]. \]

where we assumed that \( P(S_t|S_{t-1}, B, \pi) \approx Q(S_t|\pi) \) and \( P(O_t|S_{t-1}, B, \pi) \approx Q(O_t|\pi) \). The first assumption can be supported by the variational free energy (VFE) decomposition in terms of accuracy and complexity. Indeed, the VFE penalizes the divergence between \( Q(S_t|\pi) \) and \( P(S_t|S_{t-1}, B, \pi) \). The second assumption can be supported as follows:

\[
P(O_t|S_{t-1}, B, \pi) = \sum_{S_t} P(O_t, S_t|S_{t-1}, B, \pi) \\
\approx \sum_{S_t} Q(O_t, S_t|\pi) \\
= Q(O_t|\pi),
\]

assuming that the posterior \( P(O_t, S_t|S_{t-1}, B, \pi) \) can be approximated by \( Q(O_t, S_t|\pi) \). The last step relies on the linearity of expectation and the expectation of a constant, leading to the final result:

\[
G(\pi, \tau) = D_{KL}\left[ Q(O_t|\pi) \middle| P(O_t) \right] + \mathbb{E}_{Q(S_t|\pi)}[H[P(O_t|S_t)]].
\]

**Appendix D: The Simplest Generative Model**

This appendix provides the smallest generative model that can be considered as an active inference agent and aims to solve the k-armed bandit problem. As shown in Figure 19, this problem is composed of \( k \) slot machines or equivalently \( k \) actions that the agent can perform. Each machine has a different probability of producing a reward, and the agent must choose the action to perform to maximize the rewards obtained. The agent observes either a reward or a punishment after the execution of an action. Additional information related to the use of active inference in the context of the multi-arms bandit (MAB) task can be found in Markovic, Stojic, Schwobel, and Kiebel (2021), where active inference was compared to other major algorithms for solving MABs such as UCB sampling and Thompson sampling.

To solve the bandit problem using active inference, the first step is to create the generative model that encodes the agent’s beliefs of the environment. Two random variables are used for this purpose: \( O \) represents the possible outcomes and \( U \) the available actions. Furthermore, \( P(O|U) \) determines how the observation depends on the action performed by the agent, and \( P(U) \) encodes any prior preference over the available actions. More precisely, \( P(O|U) \) and \( P(U) \) are categorical distributions defined as follows:

\[
P(O = i|U = j) = A_{ij} \quad \text{and} \quad P(U = j) = a_j,
\]
Figure 19: This figure illustrates the three-armed bandit problem and the generative model used by the agent. Three slot machines are available to the agent, and each machine has a different probability of producing a reward. Additionally, there are two possible outcomes when pulling a lever: the agent either wins plenty of money or gets nothing. The generative model is composed of two nodes representing the possible outcomes and actions. Finally, the agent’s goal is to maximize the rewards obtained by picking the best strategy.

where \( A_{ij} \) defines the probability of the \( i \)th outcome given that the \( j \)th action is performed, and \( a_j \) encodes the prior over the \( j \)th action. Note that even if the active inference framework provides a way to learn the matrix \( A \), this section assumes that it is given to the agent. The next step is to pick an inference method to compute the posterior over the hidden state \( U \). This section keeps things simple and uses Bayes’ theorem:

\[
P(U = j | O = 1) = \frac{P(O = 1 | U = j) P(U = j)}{P(O = 1)} = \frac{A_{1j} a_j}{\sum_k A_{1k} a_k}
\]

where the definition of the generative model has been used in the last step and we conditioned on \( O = 1 \) to infer the action that is more likely to be rewarding. At this point, it is possible to act in our environment either by sampling the next action to perform from the posterior \( P(U | O = 1) \) or by picking the action with the highest posterior probability. Additionally, the posterior can be reused as an empirical prior for the next time step as follows:

\[
P(U = j) \leftarrow P(U = j | O = 1) = \frac{A_{1j} a_j}{\sum_k A_{1k} a_k}.
\]

This simple example does not capture the entire theoretical power of the active inference framework. Nevertheless, it illustrates four important concepts related to the design and use of an active inference agent: the design
of a generative model, the inference of the latent variable(s), the action selection process, and the use of the posterior as an empirical prior.

Appendix E: Possible Future Research

In this appendix, we propose future research directions aiming to understand the relationship between \( P(\pi | \gamma) \) and \( P(\pi | \alpha) \). The first direction relies on the following link between Dirichlet and gamma distributions. If we let \( X_1, \ldots, X_k \) be mutually independent random variables, each having a gamma distribution with parameters \( \theta_i \) for \( i = 1, \ldots, k \), and if we define \( Y_i = \frac{X_i}{\sum_{j=1}^{k} X_j} \) for \( i = 1, \ldots, k \), then \( (Y_1, \ldots, Y_k) \sim \text{Dir}(\theta_1, \ldots, \theta_k) \). This naturally leads to the hypothesis that the new generative model might be a generalization of the old generative model when all \( \theta_i \) are equal.

Another interesting fact that could be examined in more detail comes from studying the variance of the Dirichlet distribution. Recall that the variance of the random variable \( Y_i \) is given by

\[
\text{Var}[Y_i] = \frac{\bar{\theta}_i(1 - \bar{\theta}_i)}{\theta_0 + 1},
\]

where \( \bar{\theta}_i = \frac{\theta_i}{\sum_{j=1}^{k} \theta_j} \) and \( \theta_0 = \sum_{j=1}^{k} \theta_j \). If we stick to our definition of \( \theta_i \), that is, \( \theta_i = c - G_j \) with \( c = \sum_{j=1}^{k} \theta_j \), then we can study how the variance of \( Y_i \) behaves as \( c \) goes to infinity. Let us begin with

\[
\lim_{c \to +\infty} \bar{\theta}_i = \lim_{c \to +\infty} \frac{\theta_i}{\sum_{j=1}^{k} \theta_j} = \lim_{c \to +\infty} \frac{c - G_i}{\sum_{j=1}^{k} c - G_j} = \lim_{c \to +\infty} \frac{c - G_i}{kc - \sum_{j=1}^{k} G_j} = \lim_{c \to +\infty} \frac{c}{kc} = \frac{1}{k},
\]

where we note that \( G_i \) and \( \sum_{j=1}^{k} G_j \) become negligible as \( c \to +\infty \). Returning to the limit of the variance

\[
\lim_{c \to +\infty} \text{Var}[Y_i] = \lim_{c \to +\infty} \frac{\bar{\theta}_i(1 - \bar{\theta}_i)}{\theta_0 + 1} = \lim_{c \to +\infty} \frac{\bar{\theta}_i(1 - \bar{\theta}_i)}{\sum_{j=1}^{k} c - G_j + 1} = 0,
\]

where we used the fact that \( \bar{\theta}_i \) tends toward \( \frac{1}{k} \) (i.e., a constant with regard to \( c \)) and therefore the variance is only influenced by the \( c \) in the denominator, which tends toward \( +\infty \). Additionally, from the definition of the mode of the Dirichlet, we see that as \( c \to +\infty \), then the mode of the distribution tends toward the center of the simplex because the \( G_i \) becomes negligible:

\[
\lim_{c \to +\infty} m_\alpha = \left[ \frac{1}{k} \cdots \frac{1}{k} \right].
\]
Combining the behavior of the variance and the mode as $c \to +\infty$, we see that as $c$ increases, the prior becomes more and more compact around the center of the simplex. In other words, the policy selection becomes more and more stochastic as $c$ increases. This is not without recalling the role of $\gamma$ as highlighted in the caption of Figure 10.

Appendix F: Messages for $B$

In this appendix, we provide the derivation of the messages for $B$, which relies on the conjugacy between a categorical and a Dirichlet distribution. Let us start with the definition of $P(B; b)$, which is a product of Dirichlet distributions that can be written in the following form:

$$
\ln P(B; b) = \ln \prod_{i,u} P(B[u]_i, b[u]_i) = \sum_{i,u} \ln \text{Dir}(B[u]_i, b[u]_i)
$$

$$
= \sum_{i,u} \left[ \frac{b[u]_i}{\ln B[u]_i} - 1 \right] \ln B[u]_i - \ln B(b[u]_i)
$$

Logarithm of Dirichlet

$$
= \left[ \frac{b[1]_{i1}}{\ln B[1]_{i1}} - 1 \right] \ln B[1]_{i1} \ldots \left[ \frac{b[U]_{i|S|}}{\ln B[U]_{i|S|}} - 1 \right] \ln B[U]_{i|S|} - \sum_{i,u} \ln B(b[u]_i).
$$

(F.1)

where $|U|$ is the number of possible actions. Let $[a, b]$ denote all the natural numbers between $a$ and $b$ (inclusive). The random matrix $B[u]$ has one child $S_{\tau}$ for each time step $\tau \in [1, T]$, where action $u$ has been predicted by the $m$th policy and its probability mass function is given by equation 7.8. Similarly, the probability mass function of $S_{\tau-1}$ is obtained from equation 7.8 by decreasing all indexes $\tau$ by one. The first step requires us to rewrite equation 7.8 as a function of $u_\theta(B)$. This can be done by using the definition of the dot product and rearranging to obtain

$$
\ln P(S_{\tau} = k | B, S_{\tau-1} = l, \pi = m)
$$

$$
= \left[ \sum_k [\pi = k][U_{l-1}^k = 1][S_{\tau-1} = 1][S_\tau = 1] \ldots \sum_k [\pi = k][U_{l-1}^k = |U|][S_{\tau-1} = |S|][S_\tau = |S|] \right] u_\theta(B).
$$

(F.2)
These second steps aim to substitute equations F.1 and F.2 within the variational message passing equation 6.2,

$$\ln Q^*(B) = \left( \begin{bmatrix} b[1]_{11} - 1 \\ \vdots \\ b[|U||S||S|] - 1 \end{bmatrix} \cdot u_B(B) \right)$$

$$+ \sum_{\tau=1}^{T} \left( \begin{bmatrix} \sum_{k}[\pi = k][U_{\tau-1}^k = 1][S_{\tau-1} = 1][S_{\tau} = 1] \\ \vdots \\ \sum_{k}[\pi = k][U_{\tau-1}^k = |U||S_{\tau-1} = |S||S_{\tau} = |S|] \end{bmatrix} \cdot u_B(B) \right) + \text{Const},$$

where $\langle \cdot \rangle$ refers to $\langle \cdot \rangle_{Q^*}$. Note that in the above equation, $b[u]_{ij}$ are hyperparameters that can therefore be considered as constants with respect to the expectation $\langle \cdot \rangle_{Q^*}$. The third step builds on this insight by pulling the summation over time steps inside the vector, factorizing by $u_B(B)$, using the linearity of expectation, and taking the exponential of both sides to obtain

$$Q^*(B) \propto \exp \{ \mu^*_B \cdot u_B(B) \}$$

$$\mu^*_B = \left( \begin{bmatrix} b[1]_{11} - 1 + \sum_{k}[\pi = k][U_{\tau-1}^k = 1][S_{\tau-1} = 1][S_{\tau} = 1] \\ \vdots \\ b[|U||S||S|] - 1 + \sum_{k}[\pi = k][U_{\tau-1}^k = |U||S_{\tau-1} = |S||S_{\tau} = |S|] \end{bmatrix} \right).$$

By looking at equations 7.8, one can see that $\langle [S_{\tau} = i] \rangle$ and $\langle [S_{\tau-1} = j] \rangle$ are the $i$th and $j$th elements of the vector $\langle u_S(S_{\tau}) \rangle$ and $\langle u_{S_{\tau-1}}(S_{\tau-1}) \rangle$, respectively. Furthermore, because $P(\pi)$ is a categorical distribution it can be expressed as

$$P(\pi | \alpha) = \left( \begin{bmatrix} \ln \alpha_1 \\ \vdots \\ \ln \alpha_{|\pi|} \end{bmatrix} \right) \cdot \left( \begin{bmatrix} [\pi = 1] \\ \vdots \\ [\pi = |\pi|] \end{bmatrix} \right),$$

(F.3)

where $|\pi|$ is the number of policies. The above equation highlights that $\langle [\pi = k] \rangle$ is the $k$th element of $\langle u_{\pi}(\pi) \rangle$. Using those three insights, we proceed with the following reparameterization (i.e., the fourth step):
where we focused on the optimal parameters because the rest remains unchanged. The last step consists of computing the expectation of $\langle u_{S_{t-1}}(S_{t} - 1)\rangle$, $\langle u_{S_{t}}(S_{t})\rangle$, and $\langle u_{\pi}(\pi)\rangle_k$ for all $i$, $j$, and $k$:

- $\langle u_{S_{t-1}}(S_{t} - 1)\rangle_i = \langle S_{t} - 1 = i \rangle = D_{t-1|i}$
- $\langle u_{S_{t}}(S_{t})\rangle_j = \langle S_{t} = j \rangle = D_{t|j}$
- $\langle u_{\pi}(\pi)\rangle_k = \langle \pi = k \rangle = \tilde{\alpha}_k$.

One last thing we need to look at is the interaction between the summation and the indicator function in the $i$th line of equation F.4. Indeed, the sum iterates over all time steps $\tau$ and all policies $k$, but the indicator function filters out all elements where the $k$th policy does not predict the $i$th action at time $\tau - 1$. Building on this insight, we can now substitute the above results in equation F.4:

$$Q^*(B) \propto \exp \left\{ \left[ b[1]_{11} - 1 + \sum_{(k,\tau) \in \Omega_1} \tilde{\alpha}_{k} \tilde{D}_{1|\pi_{\tau} = 1} \right] \cdot u_B(B) \right\}.$$ 

Finally, one can recognize in the above equation the logarithm of a product of Dirichlet distributions written into their exponential form, that is,

$$Q^*(B) = \prod_{u,i} \text{Dir}(B[u]_{i}, b[u]_{i}) \quad \text{where} \quad b[u] = b[u] + \sum_{(k,\tau) \in \Omega_2} \tilde{\alpha}_{k} \tilde{D}_{\tau} \otimes \tilde{D}_{\tau-1}.$$ 

### Appendix G: Messages for $S_{\tau}$

This appendix shows how to derive the messages for $S_{\tau}$ for all time steps. We use equations 7.2 and 7.8, which describe $P(S_{\tau}|D)$ and $P(S_{\tau}|S_{\tau-1}, B, \pi)$ as a function of $u_{S_{\tau}}(S_{\tau})$. The first step requires us to rearrange equation 7.7 and $P(S_{\tau+1}|S_{\tau}, B, \pi)$ as a functions of $u_{S_{\tau}}(S_{\tau})$, where $P(S_{\tau+1}|S_{\tau}, B, \pi)$ is obtained by adding one to all instances of $\pi$ in equation 7.8. Those two rearrangements lead to the following results:
\[
\ln P(O_t = k| A, S_t = l) = \left[ \sum_i [O_t = i] \ln A_{il} \right] \cdot \left[ S_t = 1 \right]. \tag{G.1}
\]

\[
\ln P(S_{t+1} = k| B, S_t = l, \pi = m) = \left[ \sum_{j,k,u} [\pi = k][U_t^k = u][S_{t+1} = j] \ln B[u]_{ij} \right] \cdot u_{S_t}(S_t). \tag{G.2}
\]

For the second step, we need to substitute equations 7.2, 7.8, G.1, and G.2 into the variational message passing equation. If \( \tau = 0 \), the parent message will come from the prior, equation 7.2, otherwise from the past, equation 7.8. Also, for all time steps such that \( \tau \leq t \), there is a message from the likelihood mapping, equation G.1, and for all time steps except \( \tau = T \), there is a message from the future, equation G.2. Putting everything together we obtain:

\[
\ln Q^*(S_t) = \begin{cases} 
[\tau = 0] \left[ \ln D_1 \right] \cdot u_{S_t}(S_t) \\
+ [\tau \neq 0] \left[ \sum_{j,k,u} [\pi = k][U_{\tau-1}^k = u][S_{\tau-1} = j] \ln B[u]_{ij} \right] \cdot u_{S_t}(S_t) \\
+ [\tau \leq t] \left[ \sum_i [O_{\tau} = i] \ln A_{il} \right] \cdot u_{S_t}(S_t) \\
+ [\tau \neq T] \left[ \sum_{j,k,u} [\pi = k][U_{\tau+1}^k = u][S_{\tau+1} = j] \ln B[u]_{ij} \right] \cdot u_{S_t}(S_t) + \text{Const.}
\end{cases}
\]

The third step requires us to factorize by \( u_{S_t}(S_t) \), use the linearity of expectation, and take the exponential of both sides,

\[
Q^*(S_t) \propto \exp \left\{ \begin{bmatrix} [\tau = 0] \mu^{*}_1 + [\tau \neq 0] \mu^{*}_2 + [\tau \leq t] \mu^{*}_3 + [\tau \neq T] \mu^{*}_4 \end{bmatrix} \cdot u_{S_t}(S_t) \right\}. \tag{G.3}
\]
where

\[
\begin{align*}
\mu_1^* &= \begin{bmatrix}
\langle \ln D_1 \rangle \\
\vdots \\
\langle \ln D_{|\mathcal{S}|} \rangle
\end{bmatrix}, \\
\mu_2^* &= \begin{bmatrix}
\sum_{j,k,u} [U_{t-1}^k = u]{\langle \pi = k \rangle}{\langle S_{t-1} = j \rangle}{\langle \ln B[u]_{ji} \rangle} \\
\vdots \\
\sum_{j,k,u} [U_{t-1}^k = u]{\langle \pi = k \rangle}{\langle S_{t-1} = j \rangle}{\langle \ln B[u]_{|\mathcal{S}|i} \rangle}
\end{bmatrix}, \\
\mu_3^* &= \begin{bmatrix}
\sum_i [O_t = i]{\langle \ln A_1 \rangle} \\
\vdots \\
\sum_i [O_t = i]{\langle \ln A_{|\mathcal{S}|} \rangle}
\end{bmatrix}, \\
\mu_4^* &= \begin{bmatrix}
\sum_{j,k,u} [U_t^k = u]{\langle \pi = k \rangle}{\langle S_{t+1} = j \rangle}{\langle \ln B[u]_{1i} \rangle} \\
\vdots \\
\sum_{j,k,u} [U_t^k = u]{\langle \pi = k \rangle}{\langle S_{t+1} = j \rangle}{\langle \ln B[u]_{|\mathcal{S}|i} \rangle}
\end{bmatrix}.
\end{align*}
\]

The fourth step is the reparameterization relying on the fact that \(\langle \ln D_i \rangle\), \(\langle \pi = j \rangle\), \(\langle S_{t-1} = k \rangle\), \(\langle \ln B[i]_{ji} \rangle\), \(\langle \ln A_i \rangle\) and \(\langle S_{t+1} = k \rangle\) are elements of \(\langle u_D(D) \rangle\), \(\langle u_d(\pi) \rangle\), \(\langle u_{S_{t-1}}(S_{t-1}) \rangle\), \(\langle u_B(B) \rangle\), \(\langle u_A(A) \rangle\) and \(\langle u_{S_{t+1}}(S_{t+1}) \rangle\), respectively. Focusing on the \(\mu^*_4\) because the rest remains unchanged, the result of the reparameterization is

\[
\begin{align*}
\mu_1^* &= \begin{bmatrix}
\langle u_D(D) \rangle|_i \\
\vdots \\
\langle u_D(D) \rangle|_{|\mathcal{S}|i}
\end{bmatrix}, \\
\mu_2^* &= \begin{bmatrix}
\sum_{j,k,u} [U_{t-1}^k = u]{\langle u_d(\pi) \rangle}{\langle u_{S_{t-1}}(S_{t-1}) \rangle}{\langle u_B(B) \rangle}{u_{|\mathcal{S}|i}} \\
\vdots \\
\sum_{j,k,u} [U_{t-1}^k = u]{\langle u_d(\pi) \rangle}{\langle u_{S_{t-1}}(S_{t-1}) \rangle}{\langle u_B(B) \rangle}{u_{|\mathcal{S}|i}}
\end{bmatrix}, \\
\mu_3^* &= \begin{bmatrix}
\sum_i [O_t = i]{\langle u_A(A) \rangle}|_i \\
\vdots \\
\sum_i [O_t = i]{\langle u_A(A) \rangle}|_{|\mathcal{S}|i}
\end{bmatrix}, \\
\mu_4^* &= \begin{bmatrix}
\sum_{j,k,u} [U_t^k = u]{\langle u_d(\pi) \rangle}{\langle u_{S_{t+1}}(S_{t+1}) \rangle}{\langle u_B(B) \rangle}{u_{1i}} \\
\vdots \\
\sum_{j,k,u} [U_t^k = u]{\langle u_d(\pi) \rangle}{\langle u_{S_{t+1}}(S_{t+1}) \rangle}{\langle u_B(B) \rangle}{u_{|\mathcal{S}|i}}
\end{bmatrix}.
\end{align*}
\]

Finally, the last step consists of computing the expectations of all sufficient statistics as follows:

- \(\langle u_D(D) \rangle|_i = \langle \ln D_i \rangle = \psi(d_i) - \psi(\sum_i d_i) \equiv D_i\)
Appendix H: Derivation of the New Expected Free Energy

In this appendix, we derive the expected free energy of our new model. First, we restate the factorization of the generative model and the variational distribution:

\[
P(O_{0:T}, S_{0:T}, \pi, A, B, D, \alpha) = P(\pi|\alpha)P(\alpha|A)P(B)P(S_0|D)P(D) \\
\prod_{t=0}^{T} P(O_t|S_t, A) \prod_{t=1}^{T} P(S_t|S_{t-1}, B, \pi), \quad (H.1)
\]

\[
Q(S_{0:T}, \pi, A, B, D, \alpha) = Q(\pi)Q(A)Q(B)Q(D)Q(\alpha) \prod_{t=0}^{T} Q(S_t). \quad (H.2)
\]

Remember from appendix C that the expected free energy is defined as

\[
\mathcal{G}(\pi) = \mathbb{E}_Q \left[ \mathcal{D}_{KL} \left[ Q(X|\pi) \parallel P(O_{0:T}, X|\pi) \right] \right], \quad (H.3)
\]

where the latent variables are \( X = \{S_{0:T}, A, B, D, \alpha\} \), \( \tilde{Q} = \tilde{Q}(O_{1:T}) \overset{\Delta}{=} \prod_{t=0}^{T} \tilde{Q}(O_t) \), and \( \tilde{Q}(O_t) \overset{\Delta}{=} \sum_{S_t} \tilde{Q}(O_t, S_t) \). Now we substitute equations H.1 and H.2 into equation H.3 and simplify by removing the terms that
Realizing Active Inference in Variational Message Passing

are constant with regard to the policy $\pi$:

$$G(\pi) = \mathbb{E}_Q\left[D_{KL}\left[ Q(S_{0:T}, A, B, D, \alpha | \pi) \parallel P(O_{0:T}, S_{0:T}, A, B, D, \alpha | \pi) \right] \right]$$

$$= D_{KL}\left[ Q(A) \parallel P(A) \right] + D_{KL}\left[ Q(B) \parallel P(B) \right] + D_{KL}\left[ Q(D) \parallel P(D) \right]$$

$$+ \sum_{\tau=1}^{T} \mathbb{E}_{Q(S_{\tau-1}, B)}\left[D_{KL}\left[ Q(S_{\tau}) \parallel P(S_{\tau}|S_{\tau-1}, B, \pi) \right] \right]$$

$$- \sum_{\tau=0}^{T} \mathbb{E}_{Q(S_{\tau}, A)Q(O_{\tau+1:T})}\left[ \ln P(O_{\tau}|S_{\tau}, A) \right]$$

$$= \sum_{\tau=1}^{T} \mathbb{E}_{Q(S_{\tau-1}, B)}\left[D_{KL}\left[ Q(S_{\tau}) \parallel P(S_{\tau}|S_{\tau-1}, B, \pi) \right] \right] + C$$

$$= \sum_{\tau=1}^{T} \mathbb{E}_{Q(S_{\tau-1}, B)}\left[ -\mathbb{E}_{Q(S_{\tau})}\left[ \ln P(S_{\tau}|S_{\tau-1}, B, \pi) \right] \right] + C$$

$$= \sum_{\tau=1}^{T} \mathbb{E}_{Q(S_{\tau-1}, B)}\left[ H[P(S_{\tau}|S_{\tau-1}, B, \pi)] \right] + C,$$

where $H[\cdot]$ refer to $-\mathbb{E}_{Q(S)}[\ln P(S_{\tau}|S_{\tau-1}, B, \pi)]$ in the last equation.

Acknowledgments

We thank Karl Friston as well as the reviewers for their valuable feedback, which greatly improved the quality of this letter.

References

Berridge, K. C. (2007). The debate over dopamine’s role in reward: The case for incentive salience. Psychopharmacology, 151(3), 391–431.

Bishop, C., & Winn, J. (2003). Structured variational distributions in VIBES. In Proceedings of Artificial Intelligence and Statistics. Society for Artificial Intelligence and Statistics.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. Journal of the American Statistical Association, 112(518), 859–877.
Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1987). Occam’s razor. *Information Processing Letters*, 24(6), 377–380.

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, 16(10), 485–488.

Bowman, H., & Li, S. (2011). Cognition, concurrency theory and reverberations in the brain: In search of a calculus of communicating (recurrent) neural systems. In A. Voronkov & M. Korovina (Eds.), *Higher-Order Workshop on Automated Runtime Verification and Debugging, Easy Chair Proceedings, Festchrift celebrating Howard Barringer’s 60th Birthday*, vol. 1. EasyChair.

Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., . . . Colton, S. (2012). A survey of Monte Carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1), 1–43.

Buckley, C. L., Kim, C. S., McGregor, S., & Selh, A. K. (2017). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, 81, 55–79.

Champion, T., Bowman, H., & Grzes, M. (2021). *Branching time active inference: The theory and its generality*. Unpublished manuscript.

Cox, M., van de Laar, T., & de Vries, B. (2019). A factor graph approach to automated design of Bayesian signal processing algorithms. *Int. J. Approx. Reason.*, 104, 185–204.

Cozman, F. G. (2000). Generalizing variable elimination in Bayesian networks. In *Proc. of the Workshop on Probabilistic Reasoning in Artificial Intelligence*. https://www.ime.usp.br/~jstern/miscellanea/General/cozman00.pdf

Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., & Friston, K. (2020). Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology*, 99.

Da Costa, L., Sajid, N., Parr, T., Friston, K., & Smith, R. (2020). *The relationship between dynamic programming and active inference: The discrete, finite-horizon case*. CoRR. abs/2009.08111.

FitzGerald, T. H. B., Dolan, R. J., & Friston, K. (2015). Dopamine, reward learning, and active inference. *Frontiers in Computational Neuroscience*, 9, 136.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1), 3–71.

Forney, G. D. (2001). Codes on graphs: Normal realizations. *IEEE Transactions on Information Theory*, 47(2), 520–548.

Fountas, Z., Sajid, N., Mediano, Mediano, P., & Friston, K. (2020). Deep active inference agents using Monte-Carlo methods. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems*, 33. Red Hook, NY: Curran.

Fox, C. W., & Roberts, S. J. (2012). A tutorial on variational Bayesian inference. *Artificial Intelligence Review*, 38(2), 85–95.

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

Friston, K. (2019). A free energy principle for a particular physics. arXiv:1906.10184.

Friston, K., Da Costa, L., Hafner, D., Hesp, C., & Parr, T. (2020). Sophisticated inference. *Neural Computation*, 33, 713–763,
Realizing Active Inference in Variational Message Passing

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., Doherty, J. O., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience and Biobehavioral Reviews, 68*, 862–879.

Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., & Ondobaka, S. (2017). Active inference, curiosity and insight. *Neural Computation, 29*(10), 2633–2683.

Friston, K. J., Parr, T., & de Vries, B. (2017). The graphical brain: Belief propagation and active inference. *Network Neuroscience, 1*(4), 381–414.

Friston, K., Parr, T., & Zeidman, P. (2018). Bayesian model reduction. arXiv:1805.07092.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience, 6*(4), 187–214.

Friston, K. J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. *Neuroscience and Biobehavioral Reviews, 90*, 486–501.

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. (2013). The anatomy of choice: Active inference and agency. *Frontiers in Human Neuroscience, 7*, 598.

Gabriels, R., Cuccaro, M. L., Hill, D., Ivers, B. J., & Goldson, E. (2005). Repetitive behaviors in autism: Relationships with associated clinical features. *Research in Developmental Disabilities, 26*, pp. 169–181.

Heins, R. C., Mirza, M. B., Parr, T., Friston, K., Kagan, I., & Pooresmaeili, A. (2020). Deep active inference and scene construction. *Frontiers in Artificial Intelligence, 3*, 81.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research, 49*(10), 1295–1306.

Kojima, M., & Kangawa, K. (2005). Ghrelin: Structure and function. *Physiological Reviews, 85*(2), 495–522.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models* Cambridge, MA: MIT Press.

Kschischang, F. R., Frey, B. J., & Loeliger, H. (2001). Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory, 47*(2), 498–519.

Lam, K. S. (2007). The repetitive behavior scale-revised: Independent validation in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*, 855–866.

Lample, G., & Charlot, D. S. (2016). Playing FPS games with deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence, 31*(1).

LeCun, Y., & Cortes, C. (2010). MNIST handwritten digit database. https://www.ibi.bionomy.org/bibtex/2935bad99fa1f6e03c25b315aa3c1032/slicsid

Levine, S. (2018). Reinforcement learning and control as probabilistic inference: Tutorial and review. arXiv:1805.09099.

Lin, W., Hubacher, N., & Khan, M. E. (2018). Variational message passing with structured inference networks. In *Proceedings of the International Conference on Learning Representations*.

Markovic, D., Stojic, H., Schwoebel, S., & Kiebel, S. J. (2021). An empirical evaluation of active inference in multi-armed bandits. arXiv:2101.08699.

Millidge, B., Tschantz, A., & Buckley, C. L. (2020). Whence the expected free energy? *Neural Computation, 33*, 447–482.
Mirza, M. B., Adams, R. A., Mathys, C. D., & Friston, K. J. (2016). Scene construction, visual foraging, and active inference. Frontiers in Computational Neuroscience, 10, 56.

Mirza, M. B., Adams, R. A., Mathys, C., & Friston, K. J. (2018). Human visual exploration reduces uncertainty about the sensed world. PLoS One, 13(1), 1–20.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing Atari with deep reinforcement learning. CoRR. abs/1312.5602.

Murphy, K., Weiss, Y., & Jordan, M. I. (2013). Loopy belief propagation for approximate inference: An empirical study. arXiv:1301.6725.

Ognibene, D., & Baldassare, G. (2015). Ecological active vision: Four bioinspired principles to integrate bottom–up and adaptive top–down attention tested with a simple camera-arm robot. IEEE Transactions on Autonomous Mental Development, 7(1), 3–25.

Parr, T., Da Costa, L., & Friston, K. (2020). Markov blankets, information geometry and stochastic thermodynamics. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375(2016), 20190159.

Parr, T., Dimitrije, M., Kiebel, S. J., & Friston, K. J. (2019). Neuronal message passing using mean-field, Bethe, and marginal approximations. Scientific Reports, 9(1).

Parr, T., & Friston, K. J. (2018). Generalised free energy and active inference: can the future cause the past? Biological Cybernetics, 113, 495–513.

Rawlik, K., Toussaint, M., & Vijayakumar, S. (2013). On stochastic optimal control and reinforcement learning by approximate inference. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (pp. 3052–3056). Stanford, CA: AAAI Press.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. Science, 275(5295), 1593–1599.

Schwartenbeck, P., Passerini, J., Hauser, T. U., FitzGerald, T. H. B., Kronbichler, M., & Friston, K. (2019). Computational mechanisms of curiosity and goal- directed exploration. eLife, 8, e4703.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Hassabis, D. (2016). Mastering the game of go with deep neural networks and tree search. Nature, 529(7587), 484–489.

Smith, R., Friston, K. J., & Whyte, C. J. (2021). A step-by-step tutorial on active inference and its application to empirical data. PsyArXiv.

Solopchuk, O. (2018). Tutorial on active inference. https://medium.com/@solopchuk/tutorial-on-active-inference-30edcf80f5dc

Sundar Rajagopalan, S., Dhall, A., & Goeree, R. (2013). Self-stimulatory behaviors in the wild for autism diagnosis. In Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway, NJ: IEEE.

Tschantz, A., Baltieri, M., Seth, A. K., & Buckley, C. L. (2020). Scaling active inference. In Proceedings of the 2020 International Joint Conference on Neural Networks (pp. 1–8). Piscataway, NJ: IEEE.

Ueltzhöffer, K. (2018). Deep active inference. Biological Cybernetics, 112(6), 547–573.

van de Laar, T., & de Vries, B. (2019a). Simulating active inference processes by message passing. Front. Robotics and AI, 6, 20.
Realizing Active Inference in Variational Message Passing

van de Laar, T. W., & de Vries, B. (2019b). Simulating active inference processes by message passing. *Frontiers in Robotics and AI*, 6, 20.

Van de Maele, T., Verbelen, T., Çatal, O., De Boom, C., & Dhoedt, B. (2021). Active vision for robot manipulators using the free energy principle. *Frontiers in Neuro robotics*, 15, 14.

van Hasselt, H., Guez, A., & Silver, D. (2015). *Deep reinforcement learning with double Q-learning*. arXiv:1509.08481v3.

Wauthier, S. T., Çatal, O., De Boom, C., Verbelen, T., & Dhoedt, B. (2020). Sleep: Model reduction in deep active inference. In T. Verbelen, P. Lamillois, C. L. Buckley, & C. De Boom (Eds.), *Active inference* (pp. 72–83). Cham: Springer.

Wiegerinck, W. (2000). Variational approximations between mean field theory and the junction tree algorithm. In C. Boutilier & M. Goldszmidt (Eds.), *Proceedings of the 16th Conference in Uncertainty in Artificial Intelligence*, (pp. 626–633). San Mateo, CA: Morgan Kaufmann.

Winn, J., & Bishop, C. (2005). Variational message passing. *Journal of Machine Learning Research*, 6, 661–694.

Xing, E. P., Jordan, M. I., & Russell, S. J. (2012). A generalized mean field algorithm for variational inference in exponential families. *CoRR*. abs/1212.2512.

Yedidia, J. S. (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Trans. Information Theory*, 51(7), 2282–2312.

Yedidia, J. S. (2011). Message-passing algorithms for inference and optimization. *Journal of Statistical Physics*, 145(4), 860–890.

Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2000). Generalized belief propagation. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, & A. Culotta, (Eds.) *Advances in neural information processing systems*, 23 (pp. 668–674). Cambridge, MA: MIT Press.

Received March 1, 2021; accepted April 20, 2021.