# All Things Active Inference

**Jack Montgomery**

MNTJAC003

Supervisor: Prof. Jonathan Shock
University of Cape Town

September 26, 2024

**Abstract**

INSERT ABSTRACT

# Contents

# 1  Introduction

Active inference is a normative framework that characterises adaptive behaviour through the general paradigm of surprise minimisation. The foundation of this framework is the Free Energy Principle (FEP) (Friston et al., 2006), which is proposed as a unifying theory of cognition (Friston, 2010). Active inference can similarly be derived from the assumption of the persistence of biological systems (Friston et al., 2012). Beyond its theoretical grounding, the framework has found applications in diverse fields, such as signal processing in hearing aids (Podusenko et al., 2022) and machine learning (Millidge, 2021). The salient point in introducing these wide-ranging derivations and applications is to emphasise that no single perspective - be it biological, computational, or cognitive - can fully encompass the framework.

The difficulty, therefore, for any introduction to the framework is how to do justice to these many perspectives. We have already stated that active inference seeks to characterise adaptive behaviour as surprise minimisation, a description that may seem unusual compared to more conventional definitions. Typically, active inference is defined as the selective sampling of sensory observations in order to minimise *free energy*. Ultimately, these two definitions are equivalent, but the former is chosen in this report to ground the discussion in more familiar language. Although the physiological experience of "surprise" does not exactly match its statistical or physical meaning, it helps provide an intuitive understanding of how this framework ties together diverse behaviours under a unified principle. The challenge we face, then, is: how can we characterise both simple biological systems, like cells, and complex systems, like the human brain, as engaging in the same underlying processes? This is the ultimately what the first part of this report aims to motivate through what Parr et al. (2022) call the "low road" and "high road" to active inference.

Along the high road," we take the minimal assumption that a biological system persists through time. Even with this simple premise, we are committed to some fundamental concepts, such as the idea that the system is distinct from its environment. This distinction is formalized by the notion of a Markov blanketa boundary that separates the system's internal states from external states. Equipped with a Markov blanket, a system implicitly creates a probabilistic representation of its environment (referred to as external states). However, a system cannot simply replicate the dynamics of its environment without eventually dissipating into it. This is where the systems "resistance" to the second law of thermodynamics comes into play: a system must actively maintain its integrity to avoid entropy and disorder. We can formalize this resistance by modeling the sensory observations and the biological system as a random dynamical system. Through this lens, we will derive what it means for a system to be "active" or persist over time, thereby motivating the system's engagement in active inference.

Complementary to the high road, we will explore the "low road" to active inference. This path begins by understanding the brain as a predictive machine, grounded in Helmholtz's theory of perception as "unconscious inference" (Helmholtz, 1867) and its modern interpretation in the Bayesian brain hypothesis (Doya et al., 2007). We will consider how these theories extend beyond perception to include the sensorimotor system, showing how prediction and inference operate in action. Additionally, we will explore how the integration of bottom-up sensory input and top-down cognitive processes in perception ties into recent studies of neural activity, particularly in visual processing. These theories introduce us to the notion of a generative model of the environment. This generative model allows the brain to perform inference which again will allow us to motivate the brain engaging in active inference.

The ideas surrounding active inference that we have discussed so far are not immediately useful from an empirical perspective. The hypotheses themselves are not, by their nature, *falsifiable*. However, how active inference and the Free Energy Principle may manifest in organisms or the brain is a more testable proposition. In this report, we will present two prominent *process theories* that have been developed to explore this manifestation. The first of these is *predictive coding*.

Predictive coding is a theory of cortical function that proposes the brain is composed of a hierarchy

of neural activity, wherein higher levels predict the activity of the lower levels. The brains main objective, according to this theory, is to minimise prediction errors between the levels. This objective is consistent with the minimisation of surprise and, by extension, with active inference. However, classical predictive coding models often deal with static sensory stimuli, which are assumed to be independent. In contrast, real-world sensory inputs are highly correlated over time. This limitation is addressed by a more recent extension of the theory known as *temporal predictive coding*. Fundamentally, temporal predictive coding incorporates the dimension of time into the model, enabling it to deal with the problem of correlated stimuli.

Temporal predictive coding provides us with a formal algorithm for perception and learning within our agent. However, planninga crucial component of active inferenceis not included in this model. To address this, we will extend the temporal predictive coding model by incorporating corollaries and components from the active inference framework.

Our agent and model will be built by considering different cortical functions such as perception, learning, and planning. For each of these, we will define analogous algorithms within our model that connect the statistical inference to cognitive functions. To demonstrate these ideas, we will develop an agent within the *adaptive Bayesian thermostat* environment. This environment, introduced by Buckley et al. (2017), involves a simple task: the agent must infer its distance from a heat source based on noisy temperature readings. The agent is equipped with a preferred temperature, which it incorporates into its planning to decide on its actions.

The adaptive Bayesian thermostat is a simple environment designed to demonstrate the core components of predictive coding and active inference. In addition, we will test the temporal predictive coding model in the CartPole environment from the *OpenAI Gymnasium* package (Towers et al., 2024). This environment, a classic control problem in the reinforcement learning literature, involves balancing a pole on a cart. Several modifications were made to the environment, including shifting it from a Markov Decision Process (MDP) to a Partially Observable MDP (POMDP), and defining a preference distribution over the agent's observations. We present the results of the models performance in this task, along with a discussion.

Additionally, we will demonstrate how active inference theories manifest in *message passing* on a factor graph. Message passing has deep roots in theories of cortical computation (Parr et al., 2019), and it is also foundational to many efficient Bayesian inference algorithms. One such algorithm we will implement is *belief propagation*. This model will be tested in another reinforcement learning baseline, the Mountain Car environment.

By applying this framework to control problems and providing a formal introduction to the underlying theories, we aim to offer the reader both a clear understanding of the rich normative theory behind active inference and an appreciation for its wide-ranging practical applications.

## 2 Review of Related Work

The foundations of active inference lie in the Free Energy Principle (Friston et al., 2006). In this seminal work, Friston combines the Helmholtzian idea of perception with the concept of free energy minimisation from statistical physics to formulate a principle applicable to the brain. He argues that the problems of inferring the causes of sensory input and learning the causal structure of their generation can be resolved through free energy minimisation. In this context, free energy measures the difference between the probability distribution of environmental states affecting the system and the internal model encoded by the system's configuration. A system can minimise free energy in two ways: by altering its internal model, which corresponds to perception, or by changing its interaction with the environment, which corresponds to action. This framework assumes that the system encodes an implicit and probabilistic model of its environment.

The Free Energy Principle was further developed in the context of biological systems by Friston (2012), who approached the environment from the perspective of dynamical systems theory. This separates the system into internal and external states, with only external states susceptible to random fluctuations. The objective is to find the dynamics of internal states that ensure the system remains within a limited set of external states, reducing uncertainty. The proposed solution is motivated by the Hamiltonian Principle of least action, linking the minimisation of free energy to the broader framework of action in physical systems. Millidge (2019) offers a more accessible introduction to the Free Energy Principle, framing its applications to biological systems in more colloquial language. In contrast to the formalism of Friston et al. (2012), Millidge (2019) simplifies the key concepts, making them more intuitive while maintaining their core theoretical integrity. This review aims to balance the formal and intuitive descriptions of active inference, reflecting both perspectives.

The works cited above present active inference as a framework for adaptive biological behaviour, grounded in the Free Energy Principle. However, there are multiple ways to approach this framework. Parr et al. (2022) introduces the metaphor of the "high road" and "low road" to active inference. The "high road" starts with minimal assumptions essentially just the persistence of the system and derives active inference using the Markov blanket and the principle of least action. Which is similar to the perspectives on active inference we have discusses so far. This approach reflects a top-down view, building from general principles toward specific computational mechanisms.

On the other hand, the "low road" starts with the concept of the Bayesian brain and extends the Helmholtzian idea of perception as unconscious inference to planning as inference. Both forms of inference minimise the same objective: variational free energy. Parr et al. (2022) Which promotes the idea that we can adjust our beliefs to better reflect the world as well as, adjust the world to better reflect our beliefs. Using this dual-road metaphor to emphasise that there are many starting points for discussing active inference, reflecting the diversity of approaches in the literature.

The second half of Parr et al. (2022) explores the practical applications of active inference, detailing specific generative models and process theories associated with the framework. Smith et al. (2022) adopts a similarly empirical approach, linking the theoretical underpinnings of active inference to practical, implementable algorithms. Da Costa et al. (2020) focuses on discrete state-space models and provides a ground-up approach to the application of active inference, bridging the gap between theory and practical implementation. This work highlights how active inference is not just a theoretical construct but also a viable framework for modelling intelligent agents in real-world scenarios.

When discussing specific implementations, Parr et al. (2019) compares different message passing algorithms, which are proposed as neural process theories underlying active inference. Notably, marginal message passing is contrasted with variational message passing (Winn et al., 2005) and belief propagation (Pearl, 1986). Despite their differences, all these algorithms can be understood as minimsating free energy, albeit with different assumptions about the factorisation of the generative modelsuch as the mean-field approximation in variational message passing and the Bethe approximation in belief propagation.

The majority of the literature on active inference focuses on discrete-state models. However, continuous-state models have also been developed. For instance, van de Laar and de Vries (2019) present an online active inference agent using variational message passing in the mountain car environment, a classic reinforcement learning task. This agent is implemented using a Forney-style factor graph and computes the posterior distribution of its generative model over a defined inference horizon.

Message passing, though closely related to active inference, is distinct in that it represents a hypothesis about the physical implementation of active inference in the brain. As such, it is falsifiable, making it a candidate for empirical testing. Another prominent implementation of active inference is predictive coding, where neural circuits minimise prediction errors between hierarchical levels of

representation. The foundational work on predictive coding was done by Rao and Ballard (1999), who modelled visual processing as a hierarchy where higher levels predict lower-level activity and adjust based on the prediction errors.

Friston (2005) extended this work, showing that predictive coding can be understood as approximate Bayesian inference and that synaptic plasticity can be modelled as Hebbian learning, allowing the system to adapt to environmental noise and recognise patterns. Predictive coding has also been extended to handle temporal sequences of stimuli, as detailed by Friston et al. (2010), who introduced the concept of generalised filtering for non-linear state-space models.

However, the implementation of predictive coding for dynamic stimuli, particularly with generalised coordinates, poses challenges for neural modelling. In response, Millidge et al. (2024) introduced a simpler model called temporal predictive coding, which handles temporal sequences with local connectivity and Hebbian plasticity rules. This approach offers a more biologically plausible alternative to Kalman filtering and addresses the issue of hard-coding temporal dependencies.

Finally, Millidge (2019) combined predictive coding with active inference, allowing agents not only to perceive but also to act in ways that minimise free energy. This hybrid model uses a hierarchical predictive coding network to drive action selection, framing the agent's interaction with the environment as a Partially Observable Markov Decision Process (POMDP). Here, action becomes another variable in the free energy minimisation process, with the agent selecting actions that reduce the divergence between predicted and actual outcomes. This allows the agent to plan and act in a manner that aligns its future states with its goals, providing a practical application of active inference to decision-making.

# 3   Theory and Background

The background and theory section of this paper is inspired by the structure presented in Parr et al. (2022). The goal for this section is to show that from two completely different sets of assumptions or starting points, we are able to derive models and frameworks that represent active inference in their respective contexts.

To do this, we will ground ourselves in a definition of active inference that aims to be agnostic to any specific approach, and can be understood intuitively, almost independent of context. The definition we will use is:

**Definition 1** (Active Inference). *Adaptive systems will change their beliefs about the outside world, as well as change the outside world itself, to reduce the surprise of their observations.*

Crucially, this definition hinges on the notion of "surprise". While the meaning of surprise may vary across domains (as we will see in random dynamical systems versus statistical contexts), the central premise remains consistent. This foundation enables us to demonstrate the broad applicability of active inference across different frameworks and interpretations.

## 3.1   High Road to Active Inference

If a biological organism is surviving in an environment governed by the second law of thermodynamics, what must it **necessarily** be doing?

Biological systems, and in particular living organisms, are typically fragile in terms of the range of states they can occupy. For instance, the human body functions within a narrow temperature range - normally around $37°$ - with fevers pushing this to around $40°$, a relatively tight margin for survival. Additionally, living systems are constantly exposed to random fluctuations from their environment, which, due to the second law of thermodynamics, tend to push their internal states towards disorder, often beyond their viable limits (Friston, 2012). Given this inherent fragility and the random nature of environmental influences, the critical question becomes: how do biological systems adapt to resist these fluctuations and maintain their state within their viable bounds?

Before diving further into this question, it is essential to establish the distinction between the system and its environment. The persistence of any system over time implies a boundary between itself and the external world—a boundary that can be formalised as a *Markov blanket.*

### 3.1.1 Markov Blankets

It is essential that any adaptive system maintains both structural autonomy and separation from its environment. Without this, the system would merely dissolve into the environment's dynamics (Parr et al., 2022). This separation is statistically formalised by the existence of a Markov blanket (Pearl, 1988) between a system and its environment:

**Definition 2** (Markov Blanket). *A Markov blanket, b, is defined as:*

$$\mu \perp\!\!\!\perp x \mid b \iff p(\mu, x|b) = p(\mu|b)p(x|b)$$

This means that when a blanket $b$ is present, the variable $x$ is independent of $\mu$ given $b$. Said another way, this defines a conditional independence between the random variables $x$ and $\mu$.

A Markov blanket is a set of variables that mediate the interaction between a system and its environment. A specific example of a dynamic Markov blanket is illustrated in Figure 17, where external states of the environment are statistically separated from the internal states of the adaptive system. Additionally, two intermediary states - sensory and action states, known as the *blanket states* - are identified.

Thus, by assuming the existence of this Markov blanket, we assert that knowing the active and sensory states provides no further information about the external states when it comes to the internal states. This is typically interpreted as the idea that internal states cannot **directly** influence external states; instead, they affect them **indirectly** via the active states. Similarly, external states cannot directly alter internal states but can only affect them through sensory states.



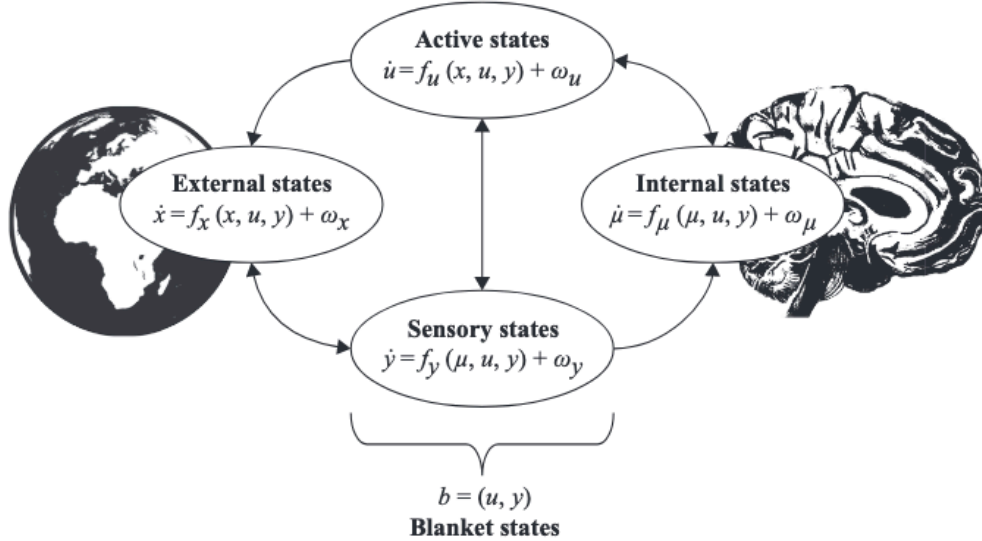Figure 1: Example of a dynamic Markov blanket from Parr et al. (2022) which separates an adaptive system, which in this case is a brain, with its environment. The dynamics of each set of states is determined by a deterministic flow specified as a function, $f$. This gives rise to the average rate of change and additional random fluctuations, $\omega$. With arrows indicating the direction of influence between the states.

The key point is that internal and external states are formally related across the Markov blanket, as both influence and are influenced by the blanket states. Because of this, we can construct conditional probability distributions for the internal and external states, given the blanket states. These distributions are conditioned on the same blanket states, allowing us to associate pairs of expected internal and external states with each other. In other words, on average, internal and external states achieve a form of synchrony. This implies that if we can define independent distributions over external and internal states given their Markov blanket, the two states become informative about each other through the blanket. This synchrony gives internal states the appearance of representing or modelling external states.
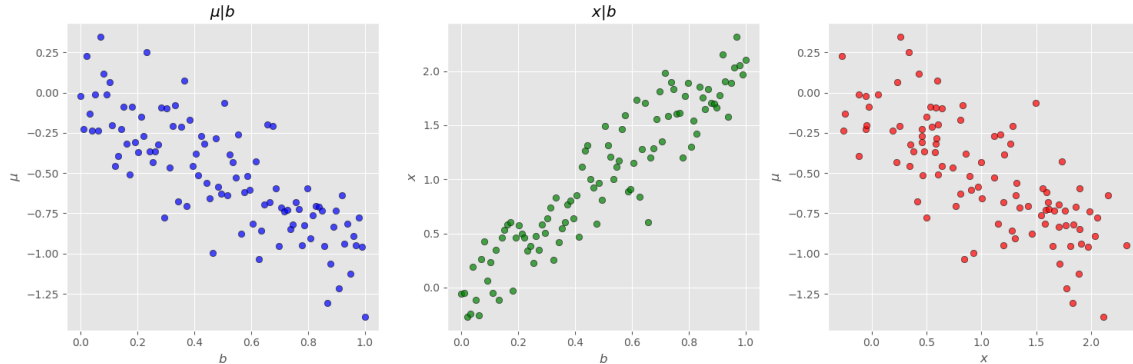


Figure 2: Example of how two random variables $x$ and $\mu$, both conditioned on the blanket state $b$ can impose some probabilistic representation of the other.

From a Bayesian perspective, the dynamics of the internal states correspond to a form of approximate Bayesian inference about external states (Parr et al., 2022). Since as sensory states change, the associated internal representation is updated through an implicit model of how sensory states are generated. However, the agents model cannot simply mimic external dynamics, as this would cause the agent to dissolve into the external environment similar to why we introduced the concept of a Markov blanket. Instead, the model must encode a preference for certain sensory states. These preferences can be specified in terms of more likely states in the models prior distribution. Under this formulation, optimal behaviour relative to the agent's prior preferences can be viewed as the maximisation of model evidence through perception and action (Parr et al., 2022). This suggests that the system engages in *self-evidencing* (Hohwy, 2016), meaning the system acts to acquire sensory states that align with its internal model, thereby maximising model evidence. We will return to these ideas later in this section.

### 3.1.2 Random dynamical system

While noting that this Markov blanket is indeed necessary for a biological system's persistence over time - the question that we posed initially is what must a system be **doing** to ensure its persistence over time. Our hypothesis for this was the system was required to keep certain internal parameters within small viable bounds. We can formalise this by describing the system and the sensory states as a random dynamical system.

A random dynamical system consists of two components: a *base flow* and a *dynamical system* on some physical space $X \in \mathbb{R}^d$.

The base flow, $\vartheta : \Omega \to \Omega$, is comprises measure-preserving measurable functions: $\vartheta_t : \Omega \to \Omega$ for each time $t \in \mathbb{R}$. These functions form a group of transformation of a probability space $(\Omega, \beta, P)$, such that $(\Omega, \beta, P, \vartheta)$ is a measure preserving dynamical system with $\sigma$-algebra $\beta$ defined on $\Omega$ and $P : \beta \to [0, 1]$ is a probability measure.

The dynamical system $(R^d, \varphi)$ comprises of a solution operator: $\varphi : \mathbb{R} \times \Omega \times X \to X$. Which is a measurable function that maps to the metric state space $X \subseteq \mathbb{R}^d$. This function satisfies the *cocycle* property: $\varphi(\tau, \vartheta_t(\omega)) \circ \varphi(t, \omega) = \varphi(t + \tau, \omega)$. This solution operator can be considered as the solution $x(t) = \varphi(t, \omega)(x_0)$ to a stochastic differential equation of the form:

$$\dot{x}(t) = f_X(x) + g_X(x)\omega(t)$$
$$x(0) = x_0$$
(1)

In all following derivations, we denote random dynamical systems as a tuple: $m = (\mathbb{R}^d, \varphi)$.

Crucially, we allow for a partition of the state space $X = R \times S$, where $R \subset X$ is distinguished by the existence of a map $\varphi_R : R \times X \to R$ that precludes direct dependency on the base flow. In this sense, $R$ constitutes an internal state space, while $S \subset X$ constitutes an external state space. Which we allow to give rise to the external mapping: $\varphi_S : \mathbb{R} \times \Omega \times A \times S \to S$. That may only depend on a subset of internal states $A \subset R$.

In this set up, internal states have dynamics that depend on external states and themselves, while external states depend on internal states and themselves but are also subject to fluctuations. See Figure 3 for a representation of these dynamics.
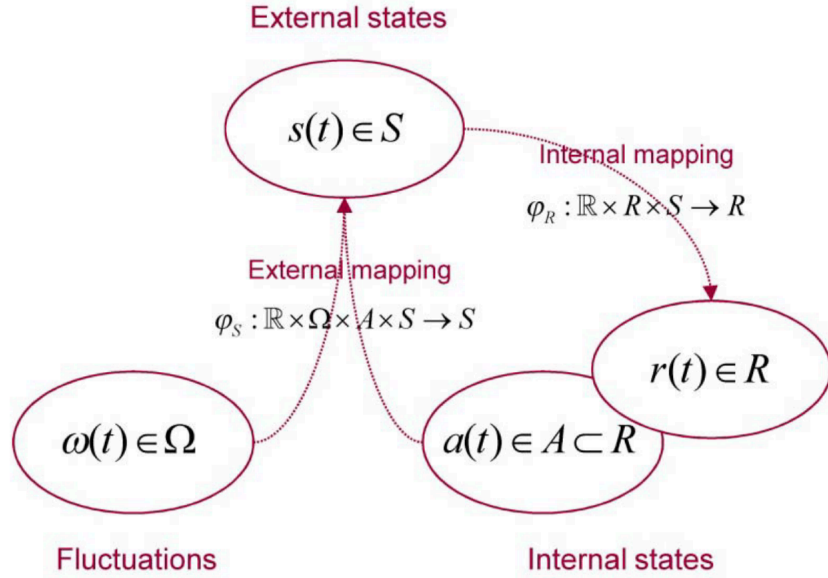


Figure 3: Image from Friston (2012) depicting the component of described random dynamical system.

**Remark**   The use of external states and internal states here is done to be consistent with Friston et al. (2012). There may be some confusion, as "external states" described in this dynamical system can be more closely associated with "sensory states" in the Markov Blanket (Figure 17).

We are interested in biological systems in this discussion thus we can regard the external states $s(t) = \varphi_S(t, \omega)(x_0)$ as the state of the sensory receptors, while the internal states $r(t) = \varphi_R(t)(x_0)$ that include $a(t) \in A \subset R$, respond to sensory perturbation. We will refer to $a(t)$ as active states since they control how environmental fluctuations are sampled by sensory states.

### 3.1.3   Active states and principle of least action

We aim to study systems whose physical states, $x(t) = \varphi(t, \omega)(x_0)$, are confined to a bounded subset of states and remain there indefinitely. This is what we have defined as necessary for a system's

persistence. Within random dynamical systems, this means the system possesses a random dynamical attractor, which we assume is ergodic. This is a compact set $\mathcal{A}(\omega) \subset X$ which. is invariant under the flow map such that $\varphi(t,\omega)(\mathcal{A}(\omega)) = \mathcal{A}(\vartheta_t(\omega))$. In terms of ergodic theory, this means there is a unique and stationary ergodic density $p(x|m)$ that is proportional to the amount of time each state is occupied. This density can then be characterised by:

$$p(s|m) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \delta(x - x(t)) \, dt \; a.s \tag{2}$$

The ergodic density $p(x|m)$ is an invariant probability measure that can be understood as the probability of finding the system in any state when observed at a random point in time. The existence of the ergodic density and its underlying attractor ensures the system has invariant characteristics that underwrite its existence over time.

**Definition 3** (Active System). *An ergodic dynamical system $m = (\mathbb{R}^d, \varphi)$ is said to be active if it possesses an internal map that satisfies the (locally) extremal condition:*

$$\varphi_R^* = argmin_{\varphi_R} H(S|m)$$
$$H(S|m) = - \int p(s|m) \ln p(s|m)$$

In other words, the internal states of an active system minimise the entropy over the external states. Which means that internal states of an active system minimise the dispersion of their external states. Which gives a formal representation of the hypothesis stated at the beginning of this section. The final link that is needed is between surprise and active systems. This link is found in the Principle of Least Action:

**Theorem 1** (Principle of Least Action). *The internal states of an active system minimise the Lagrangian or surprise, $\mathcal{L}(s(t))$, such that the variation $\delta_t \mathcal{S}$ of action $\mathcal{S}$ with respect to its internal states vanishes $r(t) \in R$ vanishes.*

$$r(t) = \varphi_R^*(t)(x_0) = \arg \min_r \mathcal{L}(s(t)) \iff \partial_r \mathcal{L}(s(t)) = 0 \iff \delta_r \mathcal{S} = 0$$
$$\mathcal{S} = \int_0^T \mathcal{L}(s(t)) \, dt$$

*Proof.* We can relate the entropy of the ergodic density $P(s(t)|m)$ in an active system using the ergodic theorem:

$$H(S|m) = - \int p(s|m) \ln p(s|m) = - \lim_{T \to \infty} \frac{1}{T} \int_0^T dt \ln p(s(t)|m) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \mathcal{L}(s(t)) \, dt$$

This means that for large enough $T$, we can specify:

$$T\mathcal{H}(S|m) = \int_0^T \mathcal{L}(s(t)) \, dt = \mathcal{S}$$

This equivalence means that $\delta_r \mathcal{S} = 0 \iff \delta_r \mathcal{H}(S|m) = 0 \iff argmin_{\varphi_R} H(S|m)$. Moreover, $\partial_r \mathcal{L}(s(t)) = 0 \iff \delta_r \mathcal{S} = 0$ by the fundamental Lemma of variational calculus. $\square$

Since we have assumed that active states, $a(t)$, are a subset of the internal states of the biological system. The Principle of Least Action can be summarised as saying that a biological system minimises the surprise of sensory states by adjusting its internal states, including active states. Which means that a persistent biological system is performing active inference.

### 3.1.4 Remark on the relation to Free Energy Principle

The previous derivations are based on Friston et al. (2012). Active inference is defined differently in this thesis. It is derived as a Corollary to the *Free Energy Principle*. The derivation of active inference in this regard bears a close relation to (approximate) Bayesian inference and variational free energy derivation of active inference that will be done along the "low road". Although, we have completed our goal to show that a persistent biological system must be conducting active inference - we will finish this section true to Friston (2012) and that definition of active inference.

### 3.1.5 Free Energy Principle and Active Inference

The fundamental issue with our current formulation is how internal states can minimise surprise given that they do not know the effect they will have on external states. We have answered **why** this is required, but we have not discussed **how** it can be done.

Intuitively, the solution regards the system as optimising a probabilistic model of the external dynamics, which is used to minimise surprise. This solution was mentioned in the previous section on Markov blanket when it was remarked that internal states create some implicit model of the external states through the conditioning on the blanket.

**Definition 4** (Generative Model). *Let $p(s(t)|m)$ be defined as in Definition 3, be expressed in terms of some arbitrary parameters $\psi(t) \in \Psi$ that are themselves random variables. We define the generative model as:*

$$p(s(t), \psi(t)) = p(s(t)|\psi(t))p(\psi(t))$$

$$p(s(t)) = \int_{\Psi} p(s(t), \psi(t)) \, d\psi$$

Where $p(s(t))$ is called the marginal likelihood.

Further, we define the proposal distribution, $q$, which plays the role of an arbitrary probability density over the parameters of the generative model. The probability density is parameterised by some subset of the internal states of the system, $\mu(t)$. Crucially, this proposal density is over the variables that parameterise the marginal likelihood of the external states. This allows us to define free energy for a biological system:

**Definition 5** (Proposal Distribution). *Let $q(\psi(t)|\mu(t))$ denote a mapping $q : R \times \Psi \times R \to R$. This serves as an arbitrary probability density defined over the parameters of the generative model called the proposal density. Further let this density be parameterised by $\mu(t) \in R$ with $\mu(t) \subset r(t)$.*

A point to note is that $\psi \in \Psi$ are not observable quantities. They are induced with a proposal density and only exist to parametersise the marginal likelihood of external states. This quantities are often referred to as *hidden states*.

**Definition 6** (Free Energy). *Free energy is now defined in terms of the proposal and generative densities as the following expectation:*

$$F(x(t)) = \int_{\Psi} q(\psi(t)|\mu(t)) \, \ln \frac{q(\psi(t)|\mu(t))}{p(s(t), \psi(t))} d\psi$$

**Theorem 2** (Free Energy Principle). *Let $m = (\mathbb{R}^d, \varphi)$ be an ergodic random dynamical system with state space $X = R \times S \in \mathbb{R}^d$. If the internal states $r(t) \in R$ minimise free energy then the system conforms to the principle of least action and is an active system. By the principle of least action (1):*

$$r(t) = \varphi_R^*(t)(x_0) = argmin_r \mathcal{F}(x(t)) \Rightarrow \delta_r \mathcal{S} = 0$$

*Proof.* We can express Free Energy as:

$$F(x(t)) = \int_\Psi q(\psi(t)|\mu(t)) \ln \frac{q(\psi(t)|\mu(t))}{p(s(t),\psi(t))} d\psi$$

$$= \int_\Psi q(\psi(t)|\mu(t)) \ln \frac{q(\psi(t)|\mu(t))}{p(\psi(t)|s(t),m)} d\psi - \ln p(s(t)|m) \qquad (3)$$

$$= \mathcal{L}(s(t)) + D_{KL}[\psi(t)|\mu(t)\|p(\psi(t)|s(t),m)]$$

By Gibbs' inequality, $D_{KL} \geq 0$, therefore free energy is minimised with respect to $\mu(t) \in R$ when the KL-Divergence is 0 (If such a solution exists). This means that $\mathcal{F}(x(t) = \mathcal{L}(s(t))$ and its variation with respect to $a(t) \in R$ vanishes:

$$a(t) = \text{argmin}_a \mathcal{F}(x(t)) \Rightarrow \partial_a \mathcal{F}(x(t)) = 0 \Rightarrow \partial_a \mathcal{L}(s(t)) = 0 \Leftrightarrow \delta_a \mathcal{S} = 0$$

From (1) we see that surprise does not depend on $\mu(t)$, which means $\partial_\mu \mathcal{L}(t) = 0 \Leftrightarrow \delta_\mu \mathcal{S} = 0$ and therefore $\delta_r \mathcal{S} = 0$. □

**Remark** In this proof, we have claimed that when minimised the KL-Divergence between the conditional distribution and proposal density is zero. This rests on the assumption that the proposal density can have the same form as the conditional density. This corresponds to *exact* Bayesian inference. When this assumption is relaxed - free energy becomes an upper bound on surprise and the inference is **approximately** Bayesian.

**Corollary 1** (Bayesian inference). *Systems that conform to the free energy principle represent the causes of their sensory states in a Bayesian sense: This follows from (3), where minimising free energy with respect to the internal states minimises the divergence between the proposal density and the posterior density over the parameters of the likelihood function:*

$$\mu(t) = \arg\min_\mu \mathcal{F}(x(t))$$

$$= \arg\min_\mu D_{KL}(q(\psi(t)|\mu(t))\|p(\psi(t)|s(t),m))$$

In other words, the optimal proposal density $q(\psi(t)|\mu(t))$ - parameterised by internal states becomes the posterior $p(\psi(t)|s(t),m)$, under the model entailed by the system.

**Corollary 2** (Active Inference). *Systems that conform to the free energy principle will selectively sample what they 'expect to see'. This follows from a final rearrangement of free energy in terms of complexity and accuracy or the log likelihood of sensory states under the proposal density:*

$$\mathcal{F}(x(t)) = \int_\Psi q(\psi(t)|\mu(t))\mathcal{L}(s(t))d\psi + \int_\Psi q(\psi(t)|\mu(t)) \ln \frac{q(\psi(t)|\mu(t))}{p(\psi(t)|s(t),m)} d\psi$$

$$= \int_\Psi q(\psi(t)|\mu(t)) \ln \frac{q(\psi(t)|\mu(t))}{p(\psi(t)|s(t),m)p(s(t)|m)} d\psi$$

$$= D_{KL}(q(\psi(t)|\mu(t))\|p(\psi(t)|m)) - \int_\Psi q(\psi(t)|\mu(t)) \ln p(s(t)|\psi(t),m)d\psi$$

$$a(t) = \arg\min_a \mathcal{F}(x(t))$$

$$= \arg\max_a \int_\Psi q(\psi(t)|\mu(t)) \ln p(s(t)|\psi(t),m)d\psi$$

This is what Friston et al. (2012) defines as active inference and manifests as a selective sampling of (typical) sensory states that are most likely under the system's posterior beliefs, by Corollary (1).

## 3.2 Low Road to Active Inference

Humans and other animals live in a world of sensory uncertainty. Through, introspection may tell us that perception is deterministic, many factors contribute to limiting the reliability of sensory information about the world. For example the mapping of 3D objects into a 2D image, neural noise introduced in early stages of sensory coding, and structural constraints on neural representations and computations (Knill and Pouget, 2004). Our brains must effectively deal with the resulting uncertainty to generate perceptual representations of the world and to guide our actions.

With aid from advances in artificial intelligent - researchers have begun to rigorously apply concepts of probability theory into perception. (Knill and Pouget, 2004) One striking observation from this work is the ways in which human observers behave as Bayesian optimal observers. A correspondence which has fundamental implications for neuroscience, particularly for neural representations of cortical functions like perception and learning. (Trommershauser et al., 2003a,b)

### 3.2.1 Hallucinations in perception

One of the earliest and perhaps most well-known investigations into hallucinations in perception was conducted by Wilder Penfield. (Penfield, 1961) Penfield suggested that the cortex is capable of producing percepts based on past experiences, even in the absence of current sensory input. He referred to the temporal lobes as the "interpretive cortex", believing that this region was responsible for interpreting the world through past experiences, based on his observations of the temporary hallucinations induced by electrical stimulation on patients. These hallucinations often included hearing familiar voices, musical tunes, or the sight of objects, which seemed to the patients like flashbacks from their past. The subjects' feedback led Penfield to believe that these memories were experienced within a brief episodic context, placed either in the present or the past. Hence, he coined the term interpretive cortex. (Knill and Pouget, 2004)

Penfield was reluctant to definitively state whether the memories themselves or merely their interpretations were elicited by the electrical stimulation of the temporal cortex. However, the most parsimonious interpretation suggests that memories of sounds, objects, or people are distributed across the cerebral cortex rather than being stored in a specific region. The medial temporal lobes, including the hippocampus, play a key role in organising these distributed memories. (Aggelopoulos, 2015) Moreover, under certain conditions, such as electrical stimulation, these memories can be as vivid as actual sensory input. When the hallucinated object and context conflicted with available sensory input, a distorted percept was reported, indicating that a process of evaluation, as proposed by Helmholtz, was at play.

Helmholtzs theory of unconscious inference posits that perception is not a mere reflection of sensory input but involves experiential data that may not be represented in the immediate stimulus. (Helmholtz, 1867) Elements not represented in the current stimulus are inferred from past experiences, leading to perception as a constructed interpretation of the world. This explains how false percepts, such as hallucinations, can occur when the brain uses past experiences to interpret incomplete or ambiguous sensory data.

### 3.2.2 Perceptual Inference

The relationship between past experiences and sensory stimuli is evident in neural activity. For example, Ishai et al. (2000) provided fMRI evidence suggesting that mental imagery of familiar objects activates subsets of the regions involved in their actual perception. Interestingly, in studies involving objects, faces, and other meaningful stimuli, activation in early visual areas (striate and peristriate cortices) did not differ significantly from that elicited by scrambled visual stimuli. This suggests that coherent, meaningful stimuli engage higher-order cortical areas responsible for integrating sensory input with prior knowledge.

Additionally, another fMRI study involving adult humans indicated that visual memories are likely stored in regions like the temporal and fusiform gyri, which are closely associated with the perception of those stimuli (Sterpenich et al., 2007). This supports the idea that perception of ecologically meaningful stimuli depends on cortical regions that are also activated during the mental imagery of those stimuli. This implies that perception is not purely driven by sensory input but also shaped by prior experiences, as the brain actively draws on memories to interpret new sensory data. Without this memory integration, certain features might remain imperceptible, highlighting that perception is an inferential process that combines bottom-up sensory input with top-down knowledge.

**Forward Models** Active inference posits that systems engage with their environments, not only through passive perception but by predicting and acting on sensory input. In this context, Miall and Wolpert (1996) introduced the concept of a forward model, explaining how the brain integrates sensory and motor information to predict the outcomes of actions. This model generates expectations of the next sensory state based on current motor commands and sensory input, serving as a crucial bridge between perception and action.

The cerebellum, traditionally linked to motor control, plays a key role in this forward modelling process. Recent research suggests that it is involved not only in motor control but also in cognitive functions, such as learning, reward processing, and the sense of agency (Welniarz et al., 2021). By integrating sensory and motor information through connections with cerebral, striatal, and spinal systems, the cerebellum allows the brain to compare predicted outcomes with actual sensory input, supporting both precise motor control and higher-level cognitive processes. This dual function reinforces the idea that motor actions, much like perception, can be understood as inferential processes aimed at minimizing uncertainty about the state of the world.

### 3.2.3 Bayesian Brain

We have explored perception as a process of inference and extended this idea to the sensorimotor system through the concept of a forward model. Now, we can deepen this discussion by looking at forward models through the lens of Bayesian inference. In active inference, the forward model is referred to as the *generative model*. This generative model describes a probability distribution over the hidden states of the environment, $x$, and the observations, $y$, which are generated by these hidden states. Formally, this is expressed as $p(x, y)$. The generative model is essential for the brains inferential processes, enabling it to predict and update beliefs about hidden states based on sensory input.

The key function of the generative model is to represent how sensory data, $y$, could be produced from hidden states, $x$. This concept is closely related to the *generative process*, which refers to the true mechanism by which sensory data is generated in the real world (see Figure 4). In this process, the actual hidden states of the environment, denoted $x^*$, produce the observed data, $y$. The brain then uses this sensory data to infer the hidden states, $x$, within its generative model. Importantly, the hidden states, $x^*$, of the real world are distinct from the hidden states, $x$, inferred by the brain. The brains generative model hypothesises possible hidden states, but these do not necessarily correspond to the true hidden states of the environment. In other words, the brains model may infer hidden states that dont actually exist, and vice versa (Parr et al., 2022).

### 3.2.4 Inference and Bayesian Optimality

So far, we have explored how the brain uses generative models to interpret and predict sensory information. These models allow the brain to integrate top-down expectations (derived from past experiences) with bottom-up sensory input to infer the hidden causes of the world. Now, we can delve deeper into how this inferential process is carried out mathematically, using principles from probability theory - specifically, Bayesian inference. The brain continuously updates its beliefs about hidden states based on incoming sensory data, which can be understood through the lens of Bayes' rule. This formalises the brain's ability to combine prior knowledge with new observations to guide perception and action.
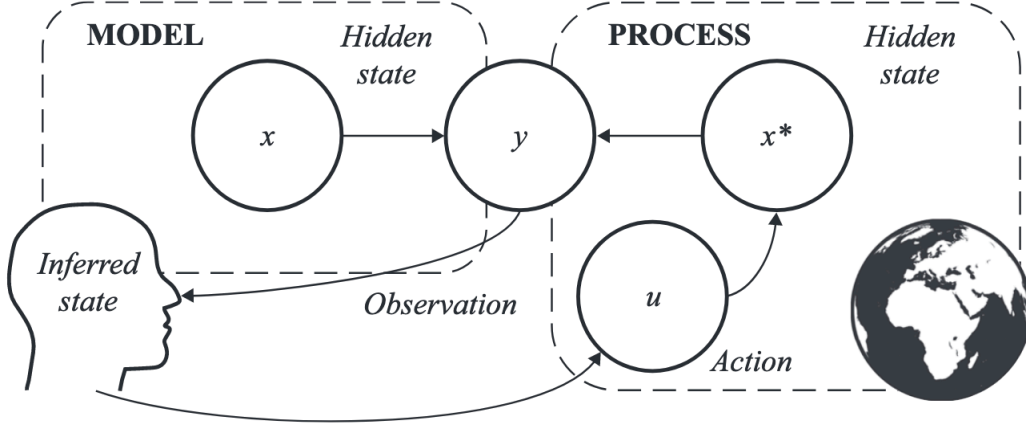
Figure 4: Image from Parr et al. (2022) depicting the generative model and process. This depictions shows there are some internal (hidden) parameters of our mode, $x$. Which are inferred by the observations generated by the environment $y$. The central distinction to note is that in theory, there exists some true state of the environment, $x^*$. This is the true causal foundation for observations in the environment. This true state may or may not be with the distribution of states specified by the prior component in the generative model - $P(x)$.

$$\underbrace{P(x|y)}_{\text{Posterior Distribution}} = \frac{P(y|x)P(x)}{P(y)} \tag{4}$$

$$\underbrace{P(x,y)}_{\text{Generative Model}} = \underbrace{P(y|x)}_{\text{Likelihood}} \underbrace{P(x)}_{\text{Prior}} \tag{5}$$

(4) shows how Bayes' rule updates our beliefs about the hidden states, $x$, given observations, $y$. The posterior distribution, $P(x|y)$, reflects the updated probability of different hidden states after accounting for the new sensory data.

(5) decomposes the generative model into two key components. The likelihood, $P(y|x)$, which expresses how likely the sensory data, $y$, is, given a specific hidden state, $x$. And prior, $P(x)$, which represents our pre-existing beliefs about the hidden states, before observing the new data.

Bayesian inference centres on obtaining the posterior distribution, which combines these two components. The term $P(y)$, known as the marginal likelihood or model evidence, serves as a normalizing constant, ensuring that the posterior is a valid probability distribution. In the next section, we will explore the significance of this term and the computational challenges it presents.

A few key points about Bayesian inference are worth noting:

1. **Integration of top-down and bottom-up dynamics**: Bayesian inference combines top-down expectations (via the prior) with bottom-up sensory evidence (via the likelihood). This integration echoes the dynamics we previously discussed in the context of perception. The equation $P(x|y) \propto P(y|x)P(x)$ captures this combination of prior knowledge and sensory input, which reflects how the brain balances stored memories and immediate data.

2. **Bayesian inference as an optimal process**: In terms of perception and action, Bayesian inference is considered optimal, in the sense that it minimises uncertainty by evaluating the full distribution over hidden states. This optimality becomes particularly important when dealing with uncertainty, as it avoids the shortcomings of approaches that rely solely on single-point estimates.

3. **Variational Free Energy (VFE)**: In the context of active inference, Bayesian optimality is achieved by minimising a quantity called variational free energy (VFE), which has roots in statistical physics (Friston et al., 2006). This measure allows the brain to manage uncertainty by aligning its beliefs with the actual sensory data, as we will explore in the next section.

It is essential to emphasise that this optimality is inherently subjective - it depends on the generative model being used. The generative model may or may not capture the true hidden states of the environment, $x^*$, assuming such states exist. In other words, the brains model of the world is always limited by its capacity to represent the actual causal dynamics of the environment. If the model fails to include important features of the environment, its predictions will be less accurate. However, Bayesian inference will still be optimal within the bounds of the chosen model. Therefore, the brains success in integrating top-down and bottom-up dynamics is constrained by the quality of its generative model.

### 3.2.5 Minimising Variational Free Energy

Having introduced the fundamentals of Bayesian inference, we now turn to two important aspects that were postponed in the previous discussion: the **marginal likelihood** and **variational free energy**. These two concepts are intimately connected and central to resolving the challenges that arise in Bayesian inference, particularly when dealing with complex generative models.

The fundamental issue with Bayesian inference lies in the computation of the marginal likelihood (or model evidence) in the expression for the posterior distribution (4). In principle, this marginal likelihood is required to normalise the posterior distribution, ensuring that it sums to 1. However, calculating it requires marginalising over all possible hidden states in the generative model:

$$p(y) = \int p(x, y) \, dx \tag{6}$$

This integral sums over the entire state space, which presents two major challenges.

Firstly, for most real-world models, this integral is too complex to solve **analytically**. The sheer complexity of generative models, especially in neural processes or high-dimensional spaces, makes exact solutions infeasible. Even when the integral can be approximated numerically, the size of the state space often makes the computation prohibitively expensive. As the dimensionality of the hidden state space increases, the computational cost of evaluating the marginal likelihood grows exponentially

To address these issues, we turn to a method known as variational inference. This approach reframes the inference problem as one of optimisation, enabling us to approximate the true Bayesian posterior in a tractable way. The key step in this process is proposing a variational posterior distributionan approximation to the true posterior that is typically chosen for its simplicity and computational ease. Inference then becomes an optimisation problem with respect to variational free energy (VFE), where the parameters of the variational posterior distribution serve as the adjustable "levers" in the optimisation process. By minimising the VFE, we find the best-fitting parameters for the variational posterior, thereby approximating the true posterior as closely as possible. Formally, variational free energy can be defined in terms of the generative model and the variational posterior as follows:

**Definition 7** (Variational Free Energy). *Let the generative model, p, and variational distribution, q, be parameterised by $\theta$ and $\phi$ respectively. Then we define variational free energy (VFE) as:*

$$\mathcal{F}(q, y) = \int q(x; \phi) \ln \frac{q(x; \phi)}{p(x, y; \theta)}$$

The claim at the end of the previous section was that Bayesian inference is (subjectively) optimal. Now, having defined variational free energy, we are indeed in a position to demonstrate this.

$$\mathcal{F}(q, y) = \int q(x; \phi) \ln \frac{q(x; \phi)}{p(x, y; \theta)}$$
$$= -\ln p(y; \theta) + \int q(x; \phi) \ln \frac{q(x; \phi)}{p(x|y; \theta)} \qquad (7)$$
$$= \underbrace{-\ln p(y; \theta)}_{\text{Evidence}} + \underbrace{D_{KL}\left[q(x; \phi) \| p(x|y; \theta)\right]}_{\text{Divergence}}$$

By Gibbs' inequality, $D_{KL} \geq 0$, further we can note that if $Q(x; \phi) = P(x|y; \theta)$ then $D_{KL} = 0$. Therefore, is the posterior distribution is the true Bayesian posterior we will be left with negative model evidence. This quantity depends solely on the generative model and not on the inferential process, indicating that Bayesian inference is optimal with respect to VFE and the generative model.

However, this does not imply that we cannot further lower the negative log evidence. In fact, one of the fundamental components of active inference is that a system actively engages with its environment. This engagement manifests in the context of VFE, by observing that a system can interact with the environment to create conditions that more closely with its beliefs, thereby minimising the VFE further. This interplay can is depicted in Figure 5.



$$F[Q, y] = \underbrace{D_{KL}[Q(x) \| P(x \mid y)]}_{\text{Divergence}} - \underbrace{\ln P(y)}_{\text{Evidence}}$$
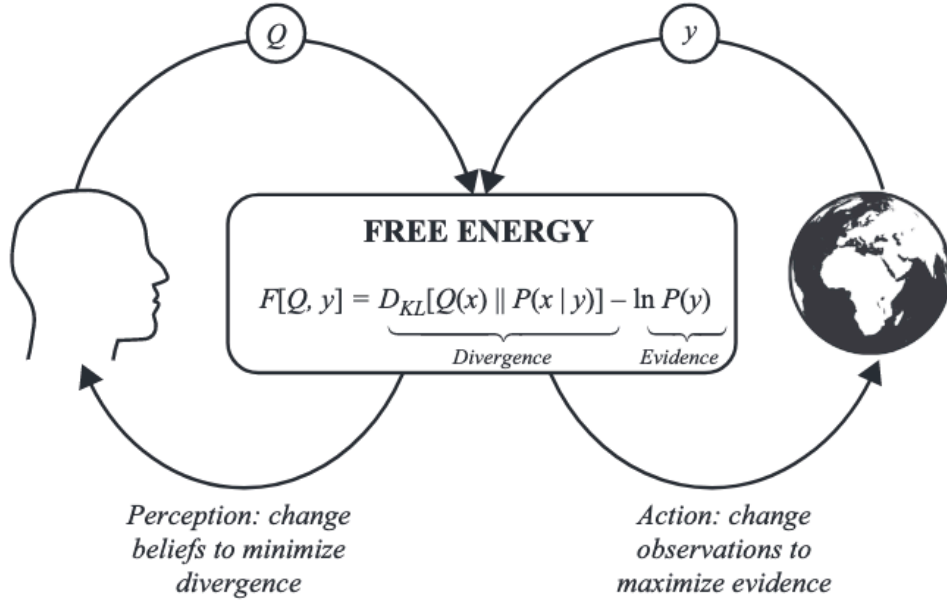
Figure 5: Image from Parr et al. (2022) depicting the interplay between a perception being a process of inference to change beliefs to better reflect the environment, and action as a process to generate observations that better reflect the system's beliefs.

The fundamental requirement for Bayesian optimality that we aimed to establish was the interplay between bottom-up and top-down inferential dynamics. These dynamics are indeed reflected in an alternative decomposition of variational free energy (VFE):

$$\mathcal{F}(Q, y) = \int Q(x; \phi) \ln \frac{Q(x; \phi)}{P(x, y; \theta)}$$
$$= \underbrace{D_{KL}\left[Q(x; \phi) \| P(x; \theta)\right]}_{\text{Complexity}} - \underbrace{\mathbb{E}_{Q(x)}\left[\ln P(y|x; \theta)\right]}_{\text{Accuracy}} \qquad (8)$$

In this decomposition, we can observe how the two terms map onto the inferential dynamics discussed earlier. The complexity term reflects the degree to which the posterior distribution diverges from the prior beliefs about the hidden states. Minimising this term requires smaller changes to prior beliefs, emphasising bottom-up influences based on past experience. In contrast, the accuracy term favours internal states that better predict the observed sensory data, thereby promoting top-down processing driven by new sensory stimuli.

### 3.2.6 Minimising Surprise

We have established the link between (approximate) Bayesian inference and top-down and bottom-up inferential dynamics required for perceptual processes. This link was achieved through the use of Variational Free Energy, however this was not the fundamental link we want to establish between Bayesian theories of the brain and active inference. We aim to link these theories to the minimisation of surprise, thus let us define this quantity in terms of the generative model:

**Definition 8** (Surprisal).
$$\mathcal{I}(y) = -\ln P(y)$$

This definition makes sense with our own notion of surprise. If we believe an observation to be 'unlikely', let us say that under our generative model $P(y) = 0.1$, then the surprise of the observation is $\mathcal{I}(y) \approx 2.3$ nats. [1] While, say $P(y) = 0.8$ then $\mathcal{I}(y) \approx 0.1$ nats.

We can notice that we have already shown that the process of Bayesian inference does indeed seek to minimise surprise:

$$\begin{aligned}
\mathcal{F}(Q, y) &= -\ln P(y; \theta) + D_{KL}\left[Q(x; \phi) \| P(x|y; \theta)\right] \\
&\geq -\ln P(y; \theta) \\
&= \mathcal{I}(y)
\end{aligned} \tag{9}$$

This demonstrates is that VFE is an upper bound on surprise. Therefore, the minimisation of VFE will result in the minimisation of surprise. Further, as we argued in the previous section, minimising VFE can be done by changing our beliefs to better reflect the environment, as well as changing the environment to reflect our beliefs. These are the conditions that we required to justify the brain performing active inference.

### 3.2.7 Expected Free Energy

While we have justified the brain's engagement in active inference, we must also recognise a fundamental *prospective* aspect of cognition that we have not mentioned as of yet: *planning*.

Within active inference, planning is similarly treated as an inferential process. To align with the Bayesian approach we have followed we must define priors and likelihoods. These distributions will be over *policies*. In active inference, a sequence of actions is referred to as a *policy*. It is important to distinguish between actions, which directly influence the external world, and policies, which represent behaviour patterns.

Future observations resulting from a particular policy are unknown, as they have not yet occurred, and must be predicted by the generative model. This prediction relies on two components: first, our belief about how hidden states evolve under a policy, denoted by $p(\tilde{x}|\pi)$, where $\tilde{x}$ represents a trajectory of hidden states. Second, the likelihood distribution, $p(y|x)$, describes the expected observations given hidden states. These two components enable the brain to simulate the counterfactual consequences of its actions, akin to asking: "What happens if I do this?" By marginalising over hidden states, we obtain the marginal likelihood (or its free energy approximation), $p(\tilde{y}|\pi)$, which reflects the evidence for a policy.

---

[1]Like bits, nats are units of information. The choice of unit depends on whether we use a logarithm to the base 2 (bits) or a natural logarithm (nats).

Active inference breaks planning into two steps: first, assigning a score to each policy, and second, forming posterior beliefs about which policy to pursue. The prior belief about which policy to follow depends on its score, where better policies are assigned higher probabilities. In active inference, the score of a policy is the negative *expected free energy*, $\mathcal{G}$, which is similar to variational free energy, except for the constraint that observations too need to be treated as random variables.

### 3.2.8 Computing EFE

We have assumed that during planning, the organism scores its policies according to their expected free energy. However, we have not yet defined what expected free energy is:

$$
\begin{align}
\mathcal{G}_\pi &= \int Q(\tilde{x}, \tilde{y}|\pi) \ln \frac{Q(\tilde{x}|\pi)}{P(\tilde{x}, \tilde{y}|\pi)} \tag{L1} \\
&= \mathbb{E}_{Q(\tilde{x}, \tilde{y}|\pi)} \left[ \ln \frac{Q(\tilde{x}|\pi)}{P(\tilde{x}, \tilde{y}|\pi)} \right] \tag{L2} \\
&\approx \mathbb{E}_{Q(\tilde{x}, \tilde{y}|\pi)} [\ln Q(\tilde{x}|\pi) - \ln Q(\tilde{x}|\tilde{y}, \pi)] - \mathbb{E}_{Q(\tilde{y}|\pi)} [\ln P(\tilde{y}|C)] \tag{L3} \\
&= - \underbrace{\mathbb{E}_{Q(\tilde{x}, \tilde{y}|\pi)} [\ln Q(\tilde{x}|\tilde{y}, \pi) - \ln Q(\tilde{x}|\pi)]}_{\text{Epistemic}} - \underbrace{\mathbb{E}_{Q(\tilde{y}|\pi)} [\ln Q(\tilde{y}|C)]}_{\text{Pragmatic}} \tag{L4}
\end{align}
$$

(10)

The definition of EFE in (L1) is identical to the definition of VFE in (7), except for the first variational posterior distribution term including the observations as well as $\tilde{x}, \tilde{y}$ representing a trajectory. The observations can then enter the expectation in (L2). This makes sense since we are considering observations that have not occurred yet. (L3) is a fundamental step in active inference, first we replace the true posterior with our variational posterior ($P(\tilde{x}|\tilde{y}, \pi) \to Q(\tilde{x}|\tilde{y}, \pi)$). Second, we drop the condition on $\pi$ in the second term, instead we condition on some distribution that is used to encode the preferences of the system, $C$. This means the system seeks to find policies expected to produce those observations. The agents preferences can be independent of the policy being followed, which allows us to drop the condition on $\pi$.

(L4) labels two terms that underscore active inference's appeal in planning. The first term is the epistemic term, notice that this term is negated, thus to minimise EFE, we need to maximise this epistemic term. What this term is representing is the difference between the variational distribution prior and conditional on an observation. That is - the agent is driven to seek out observations that reduce uncertainty about hidden states (Smith et al., 2022) The second term is the pragmatic term which as mentioned allows us to score the observations with respect to the agent's preferences.

## 4 Applications

In this section, we move from the conceptual foundation of active inference to its practical applications, aiming to demonstrate how active inference, alongside other computational neural models, can be applied to planning and control problems. This marks a shift from the previous discussion, which focused on active inference as a theoretical framework without direct empirical validation. Here, we explore the *process theories* that propose neural implementations of the computations central to active inference.

Process theories are crucial for understanding how the abstract principles of active inference can be realised within neural systems. These theories suggest that neural networks implement inference by propagating locally computed messages between neurons, constrained by the network's structure and synaptic connections (Parr et al., 2019). This imposes specific limitations on the kinds of models that can be employed. For example, traditional multi-layer perceptrons trained via backpropagation require a global error signal, which is incompatible with the localised, synapse-driven computations

hypothesised by process theories. Instead, we will focus on models like predictive coding and message-passing algorithms, which respect these constraints and offer a more biologically plausible framework for active inference.

All of the environments we consider will be framed as Markov decision processes (MDPs), a common formalism for modelling decision-making under uncertainty. Before discussing neural implementations, it is important to briefly review the key aspects of MDPs, as they provide the mathematical foundation for the problem that will be discussed later on in this section.

**Markov Decision Processes** A Markov Decision Process (MDP) is a mathematical framework used to model decision-making problems where outcomes are both under the control of the agent and subject to stochastic influences. An MDP consists of a set of states, actions, transition probabilities, and rewards. At each time step, the system occupies a particular state, and the agent selects an action. This action results in a transition to a new state, according to a probability distribution that models the systems dynamics. While classical MDPs aim to maximise expected cumulative rewards, in active inference, the objective is reframed as minimising free energy. This subtly shifts the focus from direct reward maximisation to ensuring that the agents beliefs and sensory observations remain consistent with its internal model of the world.

A key feature of MDPs is the *Markov property*, which ensures that the future state depends only on the current state and action, not on the sequence of past states or actions. This leads to a *Markov chain* over the hidden states, where the evolution of the system can be fully described by its current state and transition probabilities.

**Observations and Hidden States: MDPs vs. POMDPs** In standard MDPs, the agent has full access to the current state of the system, allowing it to make decisions based directly on this information. However, real-world environments often involve uncertainty or partial observability, which brings us to a distinction between MDPs and Partially Observable Markov Decision Processes (POMDPs).

**MDP**: In an MDP, the agent has full observability, meaning it knows the exact state of the system at each time step. Decisions are made based on the current state, and the transitions form a fully known Markov chain.

**POMDP**: By contrast, a POMDP introduces the concept of hidden statesunderlying states that are not directly observable by the agent. Instead of receiving perfect state information, the agent receives noisy or incomplete observations, which it must use to infer the true hidden state. This inference is typically performed by maintaining a belief statea probability distribution over possible hidden states. The agents decisions must therefore account not only for the underlying dynamics of the hidden states but also for the uncertainty inherent in the observations.

The distinction between MDPs and POMDPs is crucial for modelling real-world decision-making tasks. While MDPs are suited to environments where the agent can directly access the state, POMDPs are more appropriate for scenarios where the agent must infer hidden states from noisy data, such as in robotic control, navigation, or decision-making under uncertainty in biological systems. This added complexity makes POMDPs a natural fit for active inference models, as they align with the uncertainty and belief-updating processes that are central to minimising free energy.

## 4.1 Predictive Coding

We now transition our implementation of active inference into continuous state spaces by presenting an active inference agent equipped with a generative model inspired by *predictive coding*. This section will differ somewhat from other implementations in this report, adopting a tutorial-style approach similar to that presented by Bogacz (2017).

Predictive coding presents a theory of cortical function - namely, that the core function of the brain is to minimise prediction error. The intuition behind this is that the brain is comprised of a hierarchy of neurons making predictions about the activity of the neurons in the level below them. (Friston, 2008) The prediction error in this case is the difference between the neural activity predicted by the higher level compared to the actual neural activity. Over time, the hierarchy of layers encodes a range of predictions at difference scales, with fine details in local sensory data represented at the lower levels and global invariant properties of the causes of sensory data encoded at higher levels (Millidge, 2021).

These prediction errors are propagated up the hierarchy and used to improve future predictions. In predictive coding, there are two main strategies for reducing prediction errors. The first is immediate inference about the hidden states of the world, akin to perception. In this process, the prior predictions of a layer are updated by integrating residual errors from the layer below, following principles of Bayesian inference. This is performed through variational inference, where the brain iteratively updates its beliefs about hidden states by minimising free energy.

Secondly, though updating the model of the world to better explain the sensory data, which corresponds to learning. Synaptic connections between neurons, characterised by synaptic weights, are adjusted based on prediction errors. These adjustments modulate the strength of connections between neurons, allowing the brain to refine its predictions about lower-level neural activity.

### 4.1.1 Origins

The origins of predictive coding as a neuroscientific theory trace back to the 1980s and 1990s, with Rao and Ballard (1999) being one of the seminal papers introducing predictive coding as a model of neural processing. In this study, predictive coding was applied to visual processing, where a hierarchical network of neurons was exposed to natural images. The model's neurons developed simple-cell-like receptive fields, while those encoding prediction errors[2] exhibited end-stopping and other extra-classical receptive-field effects (Rao and Ballard, 1999).

Predictive coding, in its fully developed mathematical form, was later extended by Friston (2003, 2005), who linked the minimisation of prediction error to the minimization of free energy within the expectation-maximisation framework (Dempster et al., 1977). This formulation unified predictive coding with the broader concept of free energy minimisation, as we will explore in more detail in subsequent sections.

### 4.1.2 Predictive Coding as Variational Inference

The modern implementation of prediction coding can be best understood as a variational inference algorithm (Millidge, 2021). Meaning that the predictive coding algorithm defines a generative model, comprising of a prior and a likelihood distribution. In the case of predictive coding these distributions are Gaussian. Further, variational inference requires a variational posterior distribution that is used as the tractable approximation to the Bayesian posterior. Let us define these components for our model.[3]

$$
\begin{aligned}
q(x; \phi) &= \delta(x - \phi) & (A) \\
y|x &\sim N(f(x; \theta), \sigma_y^2) & (B) \\
x &\sim N(\bar{\mu}, \sigma_x^2) & (C)
\end{aligned}
\tag{11}
$$

(A) - This is our variational posterior distribution. $\delta$ denotes the Dirac-delta distribution which is a point mass probability distribution. The alternative form of the variational posterior is a Gaussian under the Laplace approximation. Friston (2003) This approximation defines the variance of a Gaussian

---

[2]The term "predictive coding" stems from the fact that certain nodes in the model explicitly encode these prediction errors

[3]We are assuming a single hidden state and a single observation dimension for this section. Hence we define the variance as a scalar value - $\sigma^2$ - , rather than as a matrix - $\Sigma$ - which will be done. Similarly, $x$, $y$ represent scalar values and not vectors.

distribution as a function of its mean. Essentially, allowing us to focus solely on estimating the first moment of the distribution. Both of these forms results in the same variational inference update equations and thus we have chosen the Dirac-delta for simplicity. Predictive coding under the Laplace approximation will be derived in Appendix B.

(B) - This is the form of our likelihood distribution. This component represents the beliefs the model has about how sensory observations, $y$, are generated from the hidden state of the model $x$. We are assuming the relation is some function $f$ of the hidden states parameterised by $\theta$. As well as variance $\sigma_y^2$

(C) - This defines the form of our prior beliefs about the hidden variable of our model, $x$. In this case, it is defined as a Gaussian with a mean of $\bar{\mu}$. We will be assuming a value of this when we operationalise this model.

The model components that are needed for inference have been defined, let us now consider an alternative decomposition of VFE that is more convenient for the derivations of the update equations in variational inference:

$$\mathcal{F} = \underbrace{\mathbb{E}_{q(x;\phi)}\left[\ln q(x;\phi)\right]}_{\text{Entropy}} - \underbrace{\mathbb{E}_{q(x;\phi)}\left[\ln p(y|x;\theta)p(x)\right]}_{\text{Energy}} \tag{12}$$

This is a useful form since we need only consider the energy term since the entropy of a Dirac-delta distribution is 0.

$$
\begin{aligned}
&\mathbb{E}_{q(x;\phi)}\left[\ln p(y|x;\theta)p(x)\right] \\
&= \int \delta(x - \phi)\ln\left(\mathcal{N}(y; f(x;\theta), \sigma_y^2)\,\mathcal{N}(x; \bar{\mu}, , \sigma_x^2)\right)dx \\
&= \ln\mathcal{N}(y; f(\phi;\theta), \sigma_y^2) + \ln\mathcal{N}(\phi; \bar{\mu}, \sigma_x^2) \\
&= -\frac{(y - f(\phi;\theta))^2}{\sigma_y^2} - \ln 2\pi\sigma_y^2 - \frac{(\phi - \bar{\mu})^2}{\sigma_x^2} - \ln 2\pi\sigma_x^2 \\
&= -\frac{1}{2}\left[\epsilon_y^2 + \epsilon_x^2 + \ln 2\pi\sigma_y^2 + \ln 2\pi\sigma_x^2\right]
\end{aligned}
\tag{13}
$$

Where, $\epsilon_y := \frac{y - f(\phi;\theta)}{\sigma_y}$ and $\epsilon_x := \frac{\phi - \bar{\mu}}{\sigma_x}$ are the prediction errors that we have reference in the predictive coding model.

This expression shows that not all errors are equal (Hodson et al., 2023). These errors are weighted by their respective standard deviation. Meaning errors with relatively higher deviation will have reduced effect on our free energy term. Anthropomorphising this, if we have a high degree of uncertainty about something we see, we tend to now take it as seriously as if we absolutely sure it is true. This is an important feature in models where the variance is treated as a model parameters to be learnt, (Bogacz, 2017; Friston, 2003) however we will not be considering this in our current model.

To operationalise the variational inference we will be taking the partial derivatives of the variational free energy. So let us put (13) and (12) together to make the optimisation objective clear:

$$-\mathcal{F} = \frac{1}{2}\left[\epsilon_y^2 + \epsilon_x^2 + \ln 2\pi\sigma_y^2 + \ln 2\pi\sigma_x^2\right] \tag{14}$$

We have negated the VFE since we aim to minimise this.

The final assumption that needs to be made is that variational free energy can be minimised using the method of gradient descent such that:

$$\frac{d\phi}{dt} = -\frac{\partial\mathcal{F}}{\partial\phi} \tag{15}$$

This key assumption governs the temporal dynamics of the hidden variable. It implies that the hidden state will evolve to minimize the prediction error over time, reflecting an essential principle of predictive coding.

Using this assumption, we can take the partial derivative of 14 with respect to the variables of interest:

$$\frac{d\phi}{dt} = -\frac{\partial \mathcal{F}}{\partial \phi} = \epsilon_y \frac{\partial f(\phi, \theta)}{\partial \phi} - \epsilon_x,$$
$$\frac{d\theta}{dt} = -\frac{\partial \mathcal{F}}{\partial \theta} = \epsilon_y \frac{\partial f(\phi, \theta)}{\partial \theta} \tag{16}$$

These update rules define our algorithm for conducting variational inference. A subtle consideration here is balancing the updates of the variational posterior parameter $\phi$ (representing the hidden state estimate) and the generative model parameter $\theta$ (representing the mapping from hidden states to observations). From a neurobiological perspective, $\phi$ corresponds to changes in neural activity, while $\theta$ reflects changes in synaptic weights. The balance between updating these variables is critical for ensuring efficient inference. From a statistical perspective, this balance can be managed using methods such as expectation maximisation.

### 4.1.3 Expectation Maximisation

Expectation-Maximisation (EM) provides a well-established procedure for the estimation of probability densities (Friston, 2003). The algorithm alternates between two steps: the Expectation (E) step and the Maximisation (M) step. In the E step, the expected values of the hidden variables are computed given the current model parameters. In the M step, the likelihood of the observed data is maximized with respect to the model parameters. Both steps are performed with respect to the same objective functionvariational free energy (Friston, 2003).

Typically, these dynamics involve fully relaxing the system with respect to the hidden variables during the E step, allowing them to reach an equilibrium. Afterward, in the M step, the system relaxes again, this time with respect to the parameters $\theta$, to maximize the likelihood. However, in our model, we take a slightly different approach. Neural activity typically operates on a much faster timescale than synaptic weight changes (Millidge et al., 2024). Thus, while we retain the full relaxation during the E step, we modify the M step by taking only a single gradient descent step on the parameters $\theta$, instead of fully relaxing the system.

There are alternative approaches to minimising variational free energy within predictive coding models. For example, Song et al. (2020) propose adjusting synaptic weights at specific intervals during the relaxation of neural activity. This approach was used in the context of approximating backpropagation in predictive coding networks. Furthermore, Salvatori et al. (2024) show that alternating gradient descent steps between the weights and the hidden states can result in both fast and stable inference in predictive coding networks. However, this was applied to large hierarchical predictive coding models which fall outside the scope of the present example. For the purposes of this report, we will adopt the standard EM-like scheme, of complete relaxation with respect to the neural activity, and then a single step with respect to the synaptic weight.

### 4.1.4 Bayesian Thermostat

We are now in a position to demonstrate the predictive coding algorithm. The environment we will be considering is the Bayesian Thermostat, as described by Buckley et al. (2017). In this environment, an agent observes the temperature from a heat source. The agents observations, denoted as $y$, correspond to temperature readings. The hidden state, $x$, represents the distance between the agent and the heat source.

We assume that the true causal structure governing the temperature readings as a function of distance is given by:

$$y = 10 - x \tag{17}$$

That is, the temperature decreases linearly as the agent moves away from the heat source. Additionally, we assume the agent receives noisy temperature readings. The relationship between temperature and distance, incorporating this noise, is modelled as:

$$y = 10 - x + \omega, \quad \omega \sim N(0, \frac{1}{25}) \tag{18}$$

Here, $\omega$ represents Gaussian noise with mean zero and variance $\frac{1}{25}$, reflecting uncertainty in the agent's temperature observations.
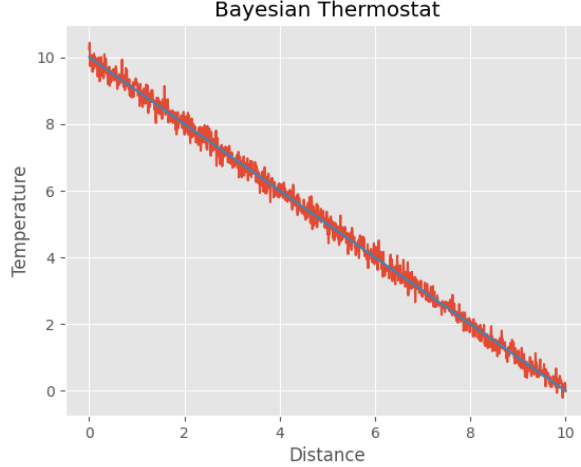


Figure 6: Visualisation of the Bayesian Thermostat environment with the true line depicting the line $y = 10 - x$. While the red is a demonstration of the type of observations the agent would receive in the environment.

### 4.1.5  Perception

The problem of perception within the Bayesian Thermostat environment is inferring the distance the agent is from the heat source, given the noisy temperature readings generated by the environment. To demonstrate this we will need to set the parameters of the generative components mentioned previously.

$$y|x \sim N\left(f(x), \frac{1}{16}\right), \quad x \sim N\left(\bar{\mu}, \frac{1}{16}\right), \quad f(x) = 10 - x \tag{19}$$

Here, $y|x$ represents the temperature readings conditioned on the hidden state $x$, and $f(x)$ captures the functional relationship between the temperature and distance.

Note the distinction between $f$, a component of the generative model, and the underlying causal structure of the data-generating process, defined earlier in (17). For the purpose of perception, we assume that $f$ accurately describes this underlying relationship. Furthermore, $f$ here is solely a function of the neural activity (or hidden state) $x$ and not the synaptic weights, as we are not yet considering learning; this will be treated in the next section.

**Remark**  Although it is formally a function of the hidden state $x$, during inference, we will treat it as a function of $\phi$, the posterior belief about $x$. This is natural since we are inferring the posterior value of $x$ through iterations on $\phi$.

The final component of our generative model is the variational posterior distribution, we will be assuming this takes the form of a Dirac-delta distribution centred at $\phi$. Thus, can now derive the update equations for the hidden variable during inference in the Bayesian Thermostat environment from (16):

$$\frac{d\phi}{dt} = -\frac{\partial \mathcal{F}}{\partial \phi} = \epsilon_y \frac{\partial f(\phi)}{\partial \phi} - \epsilon_x$$
$$= -\epsilon_y - \epsilon_x$$

(20)

During simulation we will update this hidden state using Euler's method, meaning we have some step size $\tau$ that is used to perform gradient on the VFE:

$$\phi_{k+1} = \phi_k + \tau \left( -\epsilon_y - \epsilon_x \right)$$

(21)

**Remark**  A point to note here is the different temporal scales we have introduced. $k$ denotes the time corresponding to the inference steps in the numerical implementation. Meaning that after convergence we update the hidden state for the environment time $t$. Precisely:

$$x_{t+1} = \phi_K$$

(22)

Where $K$ denotes the inference steps that were needed for convergence. Which is a hyperparameter for the model set at 10 in this case.

Suppose the agent's prior belief of its distance to the heat source, $\bar{\mu}$, is 4. Meaning that the agent is expecting to observe a temperature reading of $f(4) = 6$. While further assume that the agent actually receives a temperature reading of $y = 2$. Meaning that it is likely that the agent is further from the heat source than its prior beliefs dictate. Indeed this is reflected in a plot of the inference steps in Figure 8.
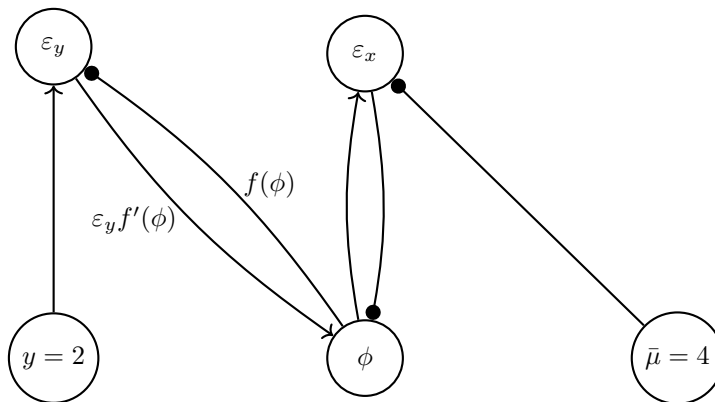


Figure 7: The architecture of the predictive coding model that is used in the perception task in the Bayesian Thermostat environment. Lines with arrows denote excitatory connections while lines ended with circles denote inhibitory connections. Lines with no labels denote a weight of 1.

What we note about the inference plot in Figure 8 is that the variational posterior distribution does indeed correctly infer the true Bayesian posterior mean. The difference between these two distribution is due to the fact that we are not learning the variance of these distributions. You can refer to Bogacz (2017); Friston (2003) for demonstrations on this. It is also interesting to note for those not familiar with Bayesian inference, how the new posterior mean of the distribution over hidden states is not the value that best describes the data, which in this case would be $x = 8$. But, the posterior beliefs represent a tradeoff between this optimal update and the priors beliefs about the hidden states.
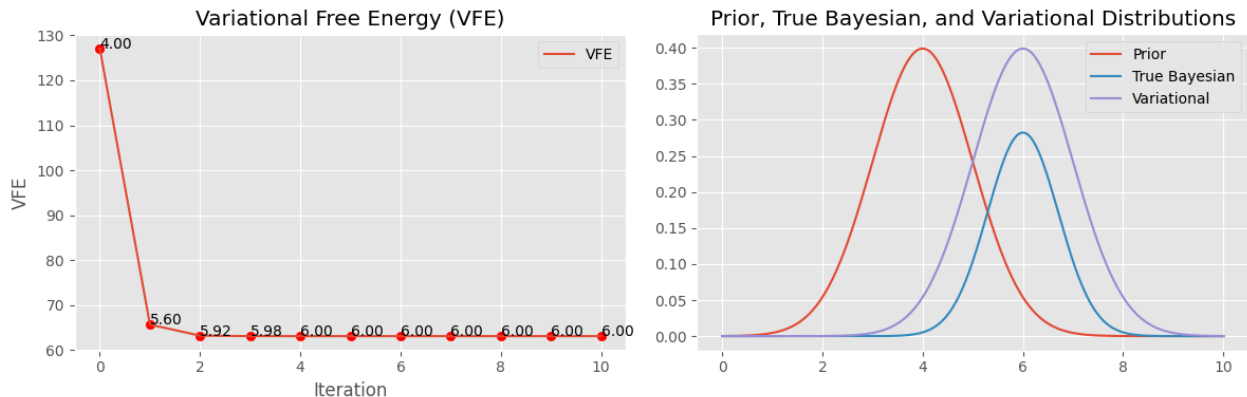
Figure 8: Left: Plot of the the values of VFE during gradient descent. Each value plotted corresponds to the value of $\phi$ used in the calculation of the VFE. Right: After convergence of the $\phi$ we then use the parameter as the mean of our prior term in the generative model. This plot shows the difference between the prior distribution, the variational posterior and the true Bayesian posterior distribution. Important to note is that we are only performing gradient descent on the $\phi$ and simply keeping the variance of the prior distribution constant after convergence.

### 4.1.6 Learning

We now aim to relax some of the assumptions that were made in the perception section. Namely, we will not assume that the agent has a perfect representation of the causal dynamics of the environment ie. we will assume that $f$ does not perfectly describe the relationship between the distance and the temperature. We will assume this function has the following form:

$$f(x, \theta) = 10 - \theta x \tag{23}$$

Practically, this means the model is required to not only infer where in the environment they are, but only learn the relation between where they are and the temperature. This is indeed somewhat circular, we require the dynamics to accurately infer where we are, but we need to know where we are to accurately infer the dynamics. This is the essential problem we are trying to solve.

To learn and update the synaptic weights we will similarly use (16) to derive our update equations for $\theta$:

$$\begin{aligned} \frac{d\theta}{dt} = -\frac{\partial \mathcal{F}}{\partial \theta} &= \epsilon_y \frac{\partial f(\phi, \theta)}{\partial \theta} \\ &= \eta \theta \epsilon_y \end{aligned} \tag{24}$$

$\eta$ is the learning rate of the model, which controls the size of the update step for the synaptic weights. This is a common parameter in neural network training and is crucial for regulating the gradient descent process. The reason for constraining the update step is that the objective surface with respect to the synaptic weights is often non-linear and not well-behaved, meaning large updates could lead to instability or poor convergence. Therefore, $\eta$ helps ensure smoother and more controlled updates.

In this example, the learning rate is set to 0.01, which is a hyper-parameter that must be tuned based on the specific problem. We will use Eulers method to update the synaptic weights as follows:

$$\theta_{t+1} = \theta_t + \tau(\eta \phi \epsilon_y)$$

Notice that this temporal scale is the environment's. Meaning that we will take a single gradient descent step with respect to the weight. We will also need to adjust the algorithm for hidden state inference given the generalisation of $f$:

$$\frac{d\phi}{dt} = -\frac{\partial \mathcal{F}}{\partial \phi} = \epsilon_y \frac{\partial f(\phi, \theta)}{\partial \phi} - \epsilon_x \qquad (25)$$

$$= \theta \epsilon_y - \epsilon_x$$

We assume the same prior and observation as the previous section. Only that we are now required to initialise the $\theta = -0.75$. We see the results from the inference scheme in 9.
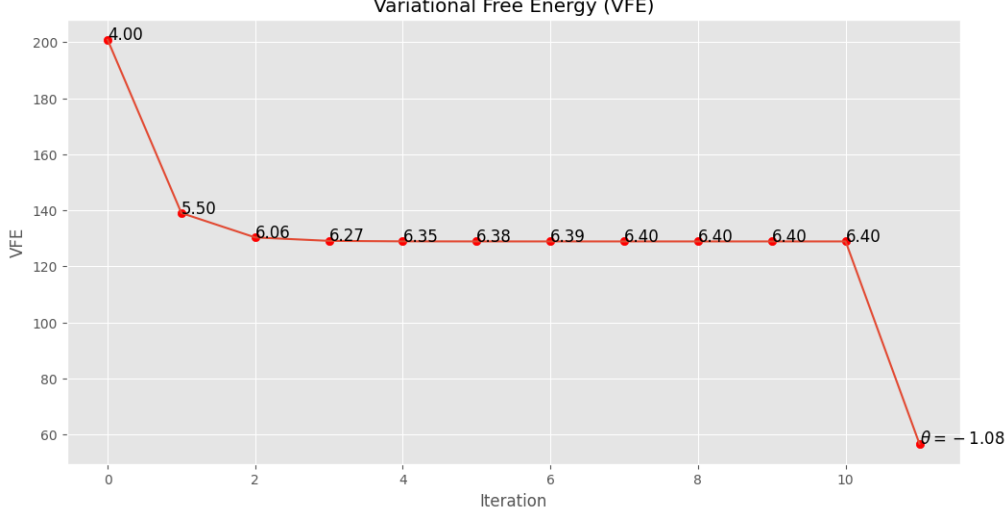


Figure 9: The gradient descent on VFE that was described in the example of learning in predictive coding. The first 10 points demonstrate the descent with respect to $\phi$ and the final step is the step with respect to $\theta$ described in (24) With hyper-parameters of $\eta = 0.01$ and $\tau = 0.01$

We observe that the variational free energy (VFE) converges to a value higher than that seen in Figure 8 in the perception section. However, the maximisation step (M-step) in the expectation-maximisation (EM) scheme brings the VFE below what was observed during perception alone. Interestingly, this update with respect to $\theta$ actually overestimates the synaptic weight (-1.08 compared to the true value of -1). This is not surprising given that the inferred hidden distance converges to 6.4. As a result, the temperature readings are better explained by a larger $\theta$ based on the models prior beliefs, leading to an overestimation of the true value.

This outcome demonstrates how inference is constrained by the model's ability to explain the environment. Since the agent is mistaken about both its distance from the heat source and its belief about the relationship between distance and temperature, it is natural to observe instability in the inference process. These inaccuracies reflect how the model's underlying assumptions can cause deviations from the true values during the learning process.

### 4.1.7   Planning

The final form of cognition in our predictive coding model is planning. Up to this point, we have focused on perception and learning without explicitly referring to active inference. However, as seen in the previous sections, variational free energy minimisation is closely related to the principles of active inference. Predictive coding, as described so far, does not inherently involve planning. To introduce planning, we will leverage the expected free energy (EFE) functional to define distributions over policies.

In this context, we will consider a single time-step into the future, which is sufficient for planning within the simplicity of this environment.

The simplification that is made in this model of planning is the transitions. We will assume that the agent knows exactly how (its beliefs) of the distance to the heat source will change with each action. We define 3 actions that the agent can take: Left, right and stay. Further, we define the transition dynamics as Dirac-delta distributions:

$$
\begin{aligned}
P(x_{t+1}|x_t, u_{t+1}^S) &= \delta(x_{t+1} - x_t) \\
P(x_{t+1}|x_t, u_{t+1}^L) &= \delta(x_{t+1} - (x_t - 0.01)) \\
P(x_{t+1}|x_t, u_{t+1}^R) &= \delta(x_{t+1} - (x_t + 0.01))
\end{aligned}
\tag{26}
$$

Where $u^L$ is the move left action, $u^S$ is stay and $u^R$ is move right. These transitions mean that the agent will move 0.01 units right when the right action is taken. Similarly, for the left action. And will stay at the sample place with the stay action.

As described in the previous section, the EFE is evaluated based on some preference distribution over actions. In this way we allow for the agent to be goal-directed. For this environment, we will again assume that the preference distribution is a Gaussian distribution over the observations:

$$
y|C \sim N(4, \frac{1}{16})
\tag{27}
$$

This preference is completely arbitrary and is defined for demonstration. But we are now able to use these components to define the expected free energy for each action. We will be choosing an alternative expression for the functional than what we saw in (10) for ease of computation:

$$
\mathcal{G}(u_{t+1}) = \underbrace{\mathbb{E}_{Q(x_{t+1}|u_{t+1})}[H[P(y_{t+1}|x_{t+1})]]}_{\text{Ambiguity}} + \underbrace{D_{KL}[Q(y_{t+1}|u_{t+1}\|P(y_{t+1}|C))]}_{\text{Risk}}
\tag{28}
$$

Let us now discuss each of these terms of how they are implemented in our scheme.

Ambiguity - This expression for EFE is particularly useful for Gaussian generative models since the entropy $(H(\cdot))$ of a normal distribution is only a function of the variance of the distribution not the mean. Since we are only considering learning of the mean this term becomes constant. Nevertheless, we have kept this term within the model since it will encourage exploration within the agent. This will become more clear when we demonstrate how exactly we create a distribution over the policies using EFE.

Risk - This is our expected divergence between the next observation and our preference distribution.

The EFE 'values' each policy we are able to undertake, however, since we are assuming planning is as inference we will want to be able to sample from a distribution over the policy values to choose which action to take. To do this we will apply the softmax $(\sigma(\cdot))$ function to convert the EFE values for each policy into a distribution:

$$
P(\mathbf{u}_{t+1}) = \sigma(-\mathcal{G}(\mathbf{u}_{t+1}))
\tag{29}
$$

The softmax function will scale the vector input of actions into a distribution of actions proportional to the negative EFE. This makes sense since our aim is to minimise free energy. This is the why we have kept in the ambiguity constant, since this will increase the EFE of all actions equally it essentially saturates the relative expected free energy of each action. Which in some instances may be detrimental to the performance of the model. However, we have found that this term allowed for more diverse actions selection initially which meant the agent learnt the dynamics of the environment in a more efficient way.

What we aim to demonstrate with planning is that the agent is able to consistently observe preferable temperature readings even in a noisy environment. What this means is that we need to dynamics

of the agent to evolve over more than just a single time-step, which has been the approach to the tutorial thus far.

We have kept prior knowledge of distance consistent with prior sections, but we have set the true distance the agent is from the heat source as $8$.[4] And allowed the agent to interact with the environment for 250 time-steps, with the results plotted in 10.
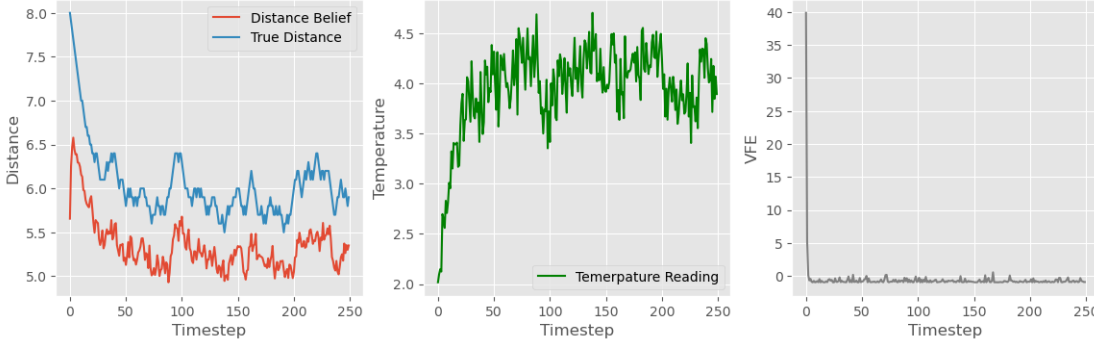


Figure 10: Results of a predictive coding agent equipped with planning using EFE in the Bayesian Thermostat environment. **Left**: This depicts the difference between the agents belief of its distance to the heat source compared to the true distance it is from the heat source. We observe that the agent is consistently estimating its distance to be greater than the true distance. **Middle**: This is a plot of the observations (temperature readings) observed by the agent over the episode. We notice that even though the agent did not have an accurate sense of where they were in the environment, they were able to consistently receive preferable observations. This is due to the fact that the dilution of the agent's own distance was offset by a weaker synaptic which in combination meant the agent converged to a reasonable niche within the environment. **Right**: This is a plot of the variational free energy after inference and learning in each time-step.

We observe that the agent consistently receives preferable temperature readings from around 50 time-steps onward and remains in this range until the end of the episode. In this sense, the agent has successfully "solved" the environment. What is particularly interesting, however, is the distorted beliefs the agent maintains about its own distance from the heat source. Throughout the episode, there is a distinct discrepancy between the agent's actual position and where it believes it is.

The reason the agent is still able to "solve" the environment despite this discrepancy lies in the range of synaptic weights, which effectively compensate for the belief distortion. The average weight over the last 200 time-steps is $-0.865$, and its average belief about its distance during the same period is 6.9. Based on the agent's beliefs about the environment's dynamics, it perceives itself to have solved the environment since $6.9 \times -0.865 + 10 = 4.03$. While this could be avoided by introducing additional components that enforce a more accurate representation of the environment, this result raises two interesting points for discussion.

The first, which we have touched on previously, is the difficulty the agent faces in jointly inferring both its location and the dynamics of the environment when there is some degree of error in one or both. The circular relationship between these elements means the agent can end up in a situation where "two wrongs make a right." In other words, errors in the model's dynamics can offset errors in its belief about its position.

The second point is related to how we interpret the hidden state $x$. We have been referring to $x$ as the agent's belief about its distance from the heat source. But in reality, it is simply a latent variable

---

[4]We have included a table of hyperparameters for this section and all previous section in Appendix A.

that the agent must infer. We have assumed the structure of the environment pushes this variable to align closely with the actual distance. However, if we relaxed certain assumptions or introduced more complex relationships between temperature and distance, this latent variable could represent something entirely different. In such a scenario, our intuition that $x$ represents distance would break down.

This leads to an important, more philosophical observation: how we, as humans, represent our environments internally may not correspond directly to the "true" states of the world. There is no guarantee that the latent representations we form are consistent across individuals. This is why our communication about the environment is grounded in shared observations, rather than in our internal states. Nevertheless, for the purposes of this report, the hidden state could represent anything, or nothing specific at all. This idea will become clearer in the next section, where we treat the size of the hidden state as a hyper-parameter.

## 4.2 Temporal Predictive Coding

In our discussion thus far we have only considered a single hidden state and single observation dimension. This model was useful for demonstration about the internal dynamics of predictive coding as well as revealing some of the complexities that arise even in simple situations with the model we are considering. We aim now to extend our understanding of predictive coding models to incorporate a deeper hierarchical structure. (Friston, 2008) We will then use the slightly more complicated formulation of the generative, and variational components of the model to define the core component for this implementation section - *temporal predictive coding*. This model of predictive coding aims to incorporate the correlation of sensory data into the components of the model. Once again, predictive coding does not include the notion of planning so we will be extending with model using EFE in a very similar way to what we have done in the previous section.

### 4.2.1 Hierarchical Predictive Coding

The success of deep neural networks in machine learning have demonstrated that having hierarchical sets of latent variables is key to enabling methods to learn powerful abstractions and to handle intrinsically hierarchical dynamics of the sort humans intuitively perceive. The predictive coding schemes previously introduced can be straightforwardly extended to handle hierarchical dynamics of arbitrary depth, equivalently to deep neural networks in machine learning.

This is done by postulating multiple layers of latent variables, $x_1, \ldots, x_L$ and defining the generative model as:

$$p(x_0, \ldots, x_L) = p(x_L) \prod_{l=0}^{L-1} p(x_l | x_{l-1}) \tag{30}$$

Where $x_l | x_{l-1} \sim N(f_l(x_{l+1}; \theta_{l+1}), \Sigma_l)$ and the final layer $x_L \sim N(\bar{x_L}, \Sigma_L)$ has some arbitrary prior $\bar{x_L}$ and the latent variable at the bottom of the hierarchy is set to what was actually observed $x_0 = y$.

Similarly, we define a variational posterior for the generative model that can be decomposed for each layer:

$$q(x_{1:L}) = \prod_{l=1}^{L} \delta(x_l - \phi_l) \tag{31}$$

The variational free energy can thus be written as the sum of prediction errors on each level:

$$\mathcal{F} = \prod_{l=l}^{L} \Sigma_l^{-1} \epsilon_l^2 + \ln 2\pi \Sigma_l \tag{32}$$

29

Where $\epsilon_l = \phi_l - f_l(\phi_{l+1}; \theta_{l+1})$[5]

Since the free energy can be nicely divided into sum of layer-wise prediction errors, it is natural to express the dynamics of $\phi$ and $\theta$ in a similar way:

$$\frac{d\phi_l}{dt} = -\frac{\partial \mathcal{F}}{d\phi_l} = \Sigma_{l-1}^{-1} \epsilon_{l-1} \frac{\partial f_{l-1}(\phi_l; \theta_l)}{\partial \phi_l} - \Sigma_l^{-1} \epsilon_l$$
$$\frac{d\theta_l}{dt} = -\frac{\partial \mathcal{F}}{d\theta_l} = \Sigma_{l-1}^{-1} \epsilon_{l-1} \frac{\partial f_{l-1}(\phi_l; \theta_l)}{\partial \theta_l} \tag{33}$$

### 4.2.2 Temporal Predictive Coding

In our examples and models of predictive coding we have mostly been focussed on treated each sensory observation independently. The problem with this approach is that sensory observations are highly correlated through time. So an essential feature of any model of the brain is how the model is able to handle this dynamically changing and correlated data. We aim to demonstrate this through a predictive coding model called temporal predictive coding (tPC).

The generative structure of the temporal predictive coding model is a hidden Markov model (HMM). This means that for every time-step $t$ there exists an observation $y_t$ that only depends on the underlying hidden state at that point in time, $x_t$. The hidden state only depends on the previous hidden state $x_{t-1}$. While this is the usually structure of a HMM, Millidge et al. (2024) compared the models performing to Kalman filter in a Bayesian filtering task, which required an additional control unit, $u_t$, which too effects the underlying state. See Figure 11 for the probabilistic graphical model of the described HMM.
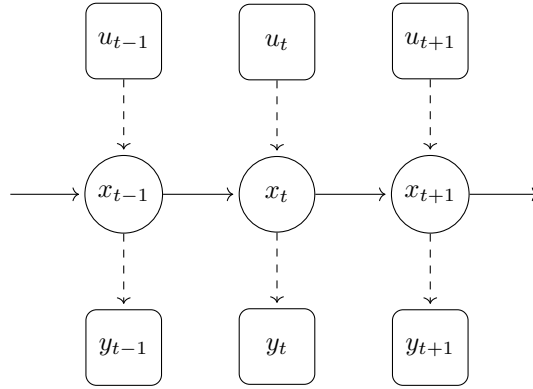


Figure 11: A Hidden Markov Model representation of the temporal predictive coding model (tPC). The model represents the recurrent connection between hidden states in the previous time step. The translations between hidden states are done by an activation on the previous level of activity, in this case this activation was a sigmoid function. The 'transition' matrix A is then applied to these activated values, further the effect of a control input $u_t$ is applied to the value of each hidden state by a multiplication with the 'control' matrix. The value of the hidden state is assumed to then generate the observations $y_t$. This emission is similarly done by passing hidden states through the activation function, with a weight matrix then applied to the output. Additionally, the emission component is equipped with a bias term in the model to scale the output to better represent the observations within the environment.

This model handles the correlation between sensory data by introducing the recurrent connection between hidden states. Machine learning and statistics have many established models with this re-

---

[5]Which is subtly different to the error in the previous section. This is simply notional, and is due to the difference between $\Sigma$ being the co-variance matrix while $\sigma$ is standard deviation.

current connection established. Which include recurrent neural networks (RNN), and more advanced versions of RNN's like Long Short-Term Memory (LSTM) models. However, for training these models use a method called backpropagation through time (BPTT). BPTT is considered to be biologically implausible because the brain processes inputs sequentially and in real time, and does not appear to have the ability to unfold sequences of computation through time. (Millidge et al., 2024)

From the graphical model of the HMM we can write out specific equations for the dynamics of the states and observations in what is called the state-space representation:

$$
\begin{aligned}
x_t &= Af(x_{t-1}) + Bu_y + \omega_x \\
y_t &= Cf(x_t) + \omega_y
\end{aligned}
\tag{34}
$$

The $A$ matrix which we will call the 'transition' matrix - encodes how the hidden states are expected to change from one time-step to the next. The $B$ encodes the effect of actions $u$ on the hidden states, which is called the 'control' matrix. Finally the $C$ is the 'emission' matrix translating the underlying state to observations[6]. While $\omega_x, \omega_y$ are Gaussian noise. We can now translate this state-space representation into our tPC generative model:

$$
\begin{aligned}
x_t|x_{t-1}, u_t &\sim N(Af(x_{t-1} + Bu_t), \Sigma_x) \\
y_t|x_t &\sim N(Cf(x_t) + c, \Sigma_y)
\end{aligned}
\tag{35}
$$

The structure of this generative model closely mirrors a typical state-space representation. The key addition is a bias term, $c$, in the emission model. This term was not included in the original formulation by Millidge et al. (2024), where the model was used exclusively for filtering tasks. In those cases, scaling the output was unnecessary. However, since we are applying this model to a control task, it becomes essential to scale the predicted output to minimise the prediction error. As discussed in the previous section, minimising prediction error is closely tied to minimising variational free energy, which remains our ultimate goal with each observation.

It is important to note that a bias term is not required in the temporal transition component of the model. This is because the hidden states in our model do not need to represent physical quantities directly. Instead, they are free to evolve in a way that is internally consistent within the model. As such, there is no need to scale them to fit external data, allowing them to serve purely as latent variables that facilitate the model's performance.

The second key component of the model is the activation function, $f$. In this case, we have chosen the sigmoid function. While the sigmoid does not have strong biological roots, other similar functions, such as the softmax, exhibit greater biophysical plausibility. Specifically, functions like the softmax reflect the dynamics of spiking neuron populations, where synaptic input is converted into neuronal firing rates in a manner consistent with observed biological processes (Deco et al., 2008).

We will be assuming the Dirac-delta distribution as our variational posterior distribution, $q = \delta(x - \phi)$. We can then define the variational free energy objective:

$$
\begin{aligned}
\mathcal{F} &= \frac{1}{2}\left(y_t - Cf\left(x_t\right) - c\right)^T \Sigma_y^{-1}\left(y_t - Cf\left(x_t\right) - c\right) \\
&+ \frac{1}{2}\left(x_t - Af\left(\hat{x}_{t-1}\right) - Bu_t\right)^T \Sigma_x^{-1}\left(x_t - Af\left(\hat{x}_{t-1}\right) - Bu_t\right)
\end{aligned}
\tag{36}
$$

Still equipped with the assumption that the components of the model change over time to minimise the variational free energy. We are able to take the partial derivative of the free energy function to define our update equations:

$$
-\frac{\partial \mathcal{F}}{\partial \phi_k} = -\epsilon_x + f'\left(\phi_k\right) \odot C^T \epsilon_y
\tag{37}
$$

---

[6]This is not the preference over observations that is traditionally also defined using 'C' in active inference.

We again note the temporal dynamics here. In the inference process we will be taking multiple gradient descent steps with respect to the posterior variational parameters. So, once $\phi$ has converged we will denote this as $\phi_K$. In which case the update is: $x_t = \phi_K$. We can take the partial derivative with respect to the other variables of interest:

$$
\begin{aligned}
\Delta A &= -\eta \frac{\partial \mathcal{F}}{\partial A} = \eta \epsilon_x f\left(\hat{x}_{t-1}\right)^T \\
\Delta B &= -\eta \frac{\partial \mathcal{F}}{\partial B} = \eta \epsilon_x u_t^T \\
\Delta C &= -\eta \frac{\partial \mathcal{F}}{\partial C} = \eta \epsilon_y f\left(x_t\right)^T \\
\Delta c &= -\eta \frac{\partial \mathcal{F}}{\partial c} = \eta \epsilon_y
\end{aligned}
\tag{38}
$$

Where $\epsilon_x := \Sigma_x^{-1}(x_t - Af(\hat{x}_{t-1}) - Bu_t)$ and $\epsilon_y := \Sigma_y^{-1}(y_t - Cf(x_t) - c)$

These updates correspond to the (M) step in the EM scheme. We will again constrain the updates of our components by a learning rate, $\eta$, the dynamics of this hyper-parameters will be discussed later in this section. What is also to note is that $\hat{x_{t-1}}$ denotes the predicted hidden state in the previous time-step, while $x_t$ denotes the relaxed neural activity from the (E) step in (37).

### 4.2.3 Extending to Planning

The extension to planning will happen in exactly the same fashion as the previous model. Meaning we will only be considering a policy of length one which means the model is only inferring what next action to perform. And similarly, we will be EFE with respect to each action and passing the negated value through the softmax function to create a distribution over actions that we then sample from to determine the next action.

The departure from the previous model is that the agent is also tasked with learning the effects of its actions on the environment. This is represented in the model through the $B$ matix. The expected next state is then expressed as:

$$
x_{t+1}|u_t \sim N(Af(\hat{x}_t) + Bu_t, \Sigma_x)
\tag{39}
$$

We will then use the mean of this distribution, $\mu_{t+1} := Af(\hat{x}_t) + Bu_t$ to create the distribution over observation in the next time-step:

$$
y_{t+1}|u_t \sim N(Cf(\mu_{t+1}) + c, \Sigma_y)
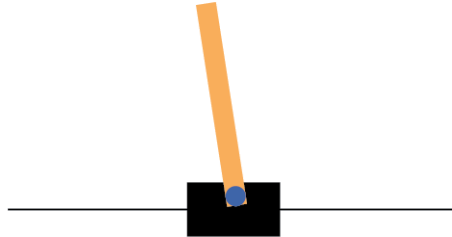\tag{40}
$$

### 4.2.4 Partially Observable Cart-Pole



Figure 12: A sample image from the OpenAI cart-pole environment

The cart-pole is a well-known environment used in reinforcement learning baselines. It requires controlling a cart with a pole attached, with the goal of balancing the pole upright. The environment

typically provides observations such as the position and velocity of the cart, as well as the angle and angular velocity of the pole. However, since velocity naturally includes some temporal information, which our model is specifically designed to capture, we modified the OpenAI (Towers et al., 2024) environment to reduce the observations to velocity only.

Furthermore, while the default environment assigns a reward at each time step, we redefined the objective to focus on maintaining the pole upright and the cart centred on the screen, similar to the approach taken by Millidge (2019). This modification fundamentally changes the control problem, as the reward feedback is now much denser. Nonetheless, this variation presents a significantly greater challenge to the predictive coding model compared to previous environment.
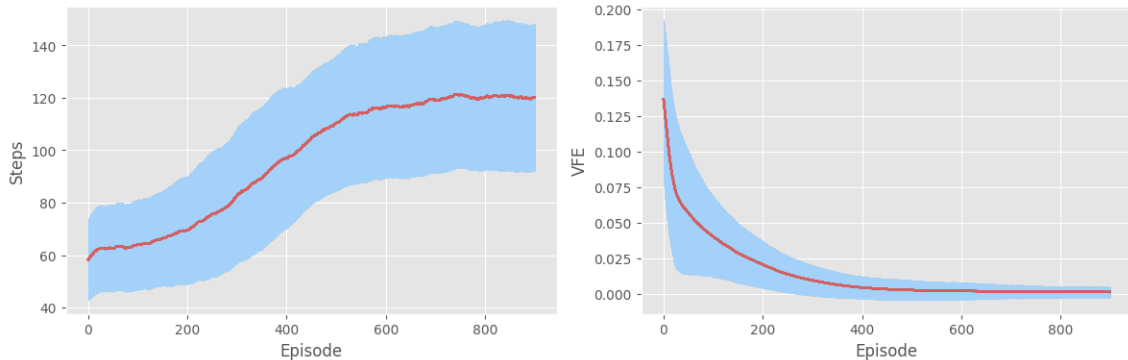
### 4.2.5    Results and Discussion



Figure 13: Results of from the partially observable cart pole environment using temporal predictive coding with planning using expected free energy. **Left**: These are the averaged smoothed episodes steps across runs. The experiment used 25 runs with each one consisting of 1000 episodes, each of these runs were then smoothed using a 100 episode moving average, these moving averages themselves were then averaged together to give the red line. The blue area depicts a confidence interval for this average. This represents a single standard deviation from the average moving average. **Right**: This depicts the associated variational free energy to the left plot. Meaning that this depicts the red line depicts the average moving average of this variational free energy per episode. The VFE per episode consisted of averaging the free energy per observation. The blue area similarly depicts a single standard deviation from the mean.

The experiment involved 1000 episodes across 25 separate runs, with the results are shown in Figure (13). he plot on the left, which illustrates the number of time steps the agent was able to balance the pole, demonstrates significant improvement after around 250 episodes, with the agent eventually maintaining balance for over 120 time steps. This indicates a non-trivial level of learning in this task. For reference, a simple baseline control rulewhere the agent moves left if the pole leansachieves an average of 50 time steps (Millidge, 2019). Additionally, a clear correlation is observable between the average variational free energy (VFE), depicted on the right, and task performance. When the model successfully reduces VFE, performance improves. Interestingly, once VFE converges near zero, performance also plateaus.

The variance in the tPC model's balance performance remains noisy, however. In some runs, the agent successfully balanced the pole for up to 500 time steps (the maximum), while in others, it dropped back to around 50 time steps. To address this, we applied exponential decay to the learning rate parameter $\eta$. While this adjustment helped the agent achieve better average performance across episodes, it did not significantly reduce the episode-to-episode variance.

A natural question arises: how well is the agent learning the dynamics of the environment? VFE provides some insight, as it acts as an upper bound on the agent's surprise given the observation, but

prediction error offers a more precise measure. If the agent is in a particular state and selects an action, part of its decision-making process involves projecting itself one step into the future a process discussed in the previous section. We can therefore compare the accuracy of these predictions by calculating the norm of the predicted observation (from the counterfactual process in the expected free energy calculation) against the actual observation. These results, shown in Figure 14, reveal that prediction errors are closely correlated with VFE, as expected. However, the model does not converge to zero prediction error, indicating it cannot fully model the environments dynamics.

This result is not unexpected. The cart-pole environments dynamics are governed by a second-order system (Florian, 2005), which poses a challenge for a single-layer network, even with a non-linear activation function.
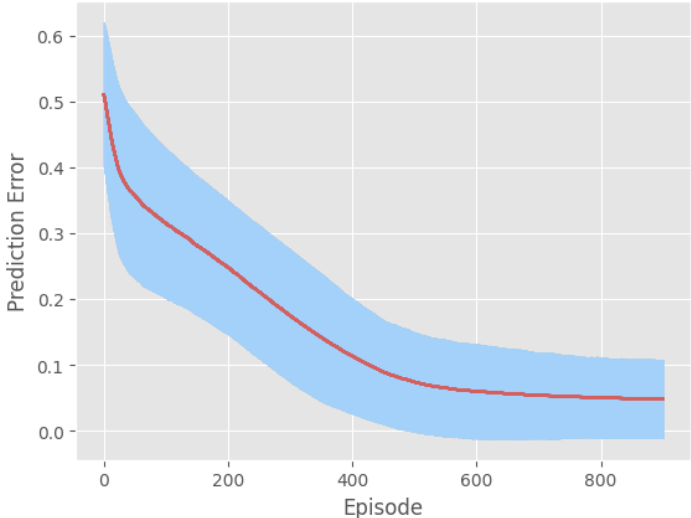


Figure 14: The averaged moving average plot for the prediction errors in the experiment. The red line depicts the average moving average from the 25 runs across episodes, while the blue shaded region depicts a single standard deviation confidence interval.

As mentioned in the previous section, we cannot directly compare the results of this experiment with other cart-pole algorithms, as the objective used in this model involves a far denser reward space. Nonetheless, Millidge et al. (2024) similarly used observations from the environment to define the preference distribution for their active inference agent. Their model, which employed a hierarchical predictive coding framework in a fully observable version of Cart-Pole, achieved an average of slightly less than 100 time steps across 50 trials.

A key factor in the performance of this model is its sensitivity to initial conditions, particularly the weight initialisation in the various matrices. In some instances, poor weight initialisation resulted in no learning across the 1000 episodes, with the agent balancing the pole for an average of only 9 time steps. To address this issue, we fixed the randomisation seed for weight initialisation across all runs in our experiments. It is important to note that this sensitivity to initialisation is not unique to this model and does not indicate an inability to scale to more complex tasks. Even deep neural networks suffer from similar weight initialisation challenges (Mishkin and Matas, 2016).

## 4.3 Message Passing

Predictive coding is often presented as a neural process theory, offering a mechanistic explanation of cognitive functions and their neuronal underpinnings. However, when constructing active inference models, predictive coding can also be understood as a variational inference algorithm for performing approximate Bayesian inference. Predictive coding is not unique in this role - many statistical inference algorithms have been suggested to have neurobiological relevance, one of which is *message passing*.

In this report, we will review and implement a specific message passing algorithm called *belief propagation*. Belief propagation allows for the exact computation of marginal posteriors, though this precision comes at the cost of greater architectural complexity compared to simpler algorithms such as variational message passing. Our focus will be on message passing over Forney-style factor graphs (FFG), also known as normal factor graphs. These graphs provide a clear graphical representation of the relationships between variables and factors within a joint probability distribution. By expressing a joint probability distribution in the form of an FFG, we can derive messages passed between nodes on the graph, enabling efficient and exact Bayesian inferenceprovided that the graph has no loops.

### 4.3.1 Forney-style Factor Graphs and Belief Propagation

A Forney-style factor graph (FFG) is a graphical representation of a factorised probabilistic model, where edges represent variables and nodes specify relations between them. The purpose of expressing a joint probability distribution as an FFG is to highlight dependencies between variables, facilitating more efficient inference. To illustrate this, consider a joint probability distribution of the form:

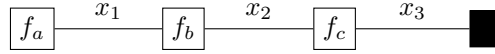$$p(x_1, x_2, x_3) = \frac{1}{Z} f_a(x_1) f_b(x_1, x_2) f_c(x_2, x_3) \tag{41}$$



Figure 15: FFG for the joint probability distribution in (41). Each factor is displayed in block, with each variable represented as an edge in the graph. The observed variable is depicted as a black node, this represents the $\delta$ function to clamp the value of the variable $x_3$.

Where $Z$ is a normalisation constant. Then let us consider the problem of inferring the marginal distribution of $x_2$ given some observation of $x_3$.

$$
\begin{aligned}
p(x_2 | x_3 = \widehat{x_3}) &\propto \int_{x_1, x_3} p(x_1, x_2, x_3) \delta(x_3 - \widehat{x_3}) \ dx_2 dx_3 \\
&\propto \int_{x_1, x_3} f_a(x_1) f_b(x_1, x_2) f_c(x_2, x_3) \delta(x_3 - \widehat{x_3}) \ dx_2 dx_3
\end{aligned}
\tag{42}
$$

We use the Dirac function, $\delta$, to clamp the variable to take the observed value.

The key point is that not all factors (functions) depend on every variable, which enables us to 'push in' the integrals. This lets us express the posterior in terms of more easily computable components, or *messages*.

$$
\begin{aligned}
p(x_2 | x_3 = \widehat{x_3}) &\propto \int_{x_1, x_3} f_a(x_1) f_b(x_1, x_2) f_c(x_2, x_3) \delta(x_3 - \widehat{x_3}) \ dx_1 dx_3 \\
&\propto \overbrace{\int_{x_1} f_a(x_1) f_b(x_1, x_2) \ dx_1}^{\mu_{f_a \to x_2}} \underbrace{\int_{x_3} f_c(x_2, x_3) \overbrace{\delta(x_3 - \widehat{x_3})}^{\mu_{x_3 \to f_c(x_3)}} \ dx_3}_{\mu_{f_c \to x_2}} \\
&\propto \mu_{f_b \to x_2} \mu_{f_c \to x_2}
\end{aligned}
\tag{43}
$$

$\mu$ is what we call a message. In the FFG we will be exploring the factors, $f$, will also be probability distributions. This means that these messages can just be considered as unnormalised probability distributions - otherwise known as beliefs. Passing these messages around the factor graph is what gives rise to the notion of belief propagation.

**Remark** The point that has not yet been address is the normalisation of these messages in order to form a valid probability distribution. This normalisation only requires the evaluation of low-dimensional integrals over local variables which makes it a more tractable computation. What is more, if all messages are Gaussian and any other node function is linear. The integrable has a closed form. (Bagaev and de Vries, 2021)

**Equality Factor** Within a FFG we depict a variable as an edge connecting two factors. As opposed to a bipartite graph where they are depicted as nodes with edges corresponding to the relationship between variables and factors. This introduces a constraint that a variable can only be a function of two factors. We resolve this constraint by an equality factor. The equality factor resolves this situation by constraining the information about three variables to be equal. (Cox et al., 2019) More precisely:

$$f_=(x, y, z) = \delta(x - z)\delta(x - y) \tag{44}$$

Meaning that:

$$\mu_{\textcircled{3}}(z) = \int\int \mu_{\textcircled{1}}(z)\mu_{\textcircled{2}}(z)\, dx_1\, dx_2$$
$$= \mu_{\textcircled{1}}(z)\mu_{\textcircled{2}}(z)$$

Figure 16: Example of the message passing behaviour over an equality factor

### 4.3.2 Mountain Car

The mountain car environment describes an environment where a car is in a valley, and the car is not powerful enough to reach its target on the side of the mountain - higher up than it currently is. Thus the car is required to move backwards and generate enough force to then move to its target on the other side of the valley.
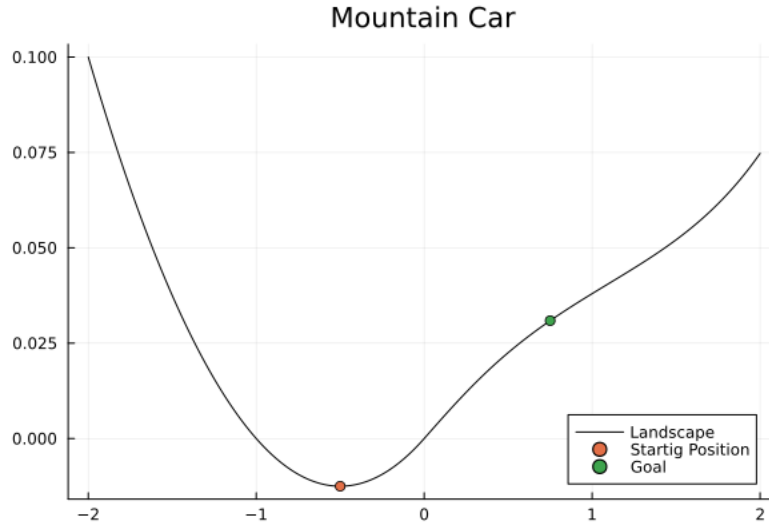
Figure 17: The landscape of the mountain car environment. Including the starting and target position for the model.

This environment was originally proposed by Moore (1990), however the specific physics that are being used are based on the environment used in Ueltzhffer (2018); van de Laar and de Vries (2019). Specifically, let us define the environmental states $z_t = (\phi_t, \dot{\phi}_t)$ depending on the position ($\phi_t$) and velocity of the car ($\dot{\phi}_t$). The environment is fully observable, meaning that the observations generated, $y_y$ , will be a reflection of these underlying states. The evolution of the state can then be described as:

$$
\begin{aligned}
\dot{\phi}_t &= \dot{\phi}_{t-1} + F_g(\dot{\phi}_{t-1}) + F_f(\dot{\phi}_{t-1}) + F_a(a_t) \\
\phi_t &= \phi_{t-1} + \dot{\phi}_t
\end{aligned}
\tag{45}
$$

Where $F_g$ is the gravitational force of the hill landscape that depends on the car's positions:

$$
F_g(\phi) = \begin{cases} -0.05(2\phi + 1), & \text{for } \phi < 0 \\ -0.05 \left[ (1 + 5\phi^2)^{-\frac{1}{2}} + \phi^2(1 + 5\phi^2)^{-\frac{3}{2}} + \frac{1}{16}\phi^4 \right] & \text{otherwise} \end{cases}
\tag{46}
$$

$F_f$ is the friction of the car are is defined as $F_f(\dot{\phi}) = -0.1\dot{\phi}$ and $F_a(a)$ is the engine force $F_a(a) = 0.04 \tanh(a)$ which restricts the force exerted by the car to be between -0.04 ad 0.04.

### 4.3.3 Model

Let us now define the form of the generative model. This model is derived from van de Laar and de Vries (2019), with some adjustment which will be highlighted. By using the Markov property of the environment we can define our generative model:

$$
p_t(x, y, u) \propto p(x_{t-1}) \prod_{k=t}^{t+T} p(y_k|x_k)p(x_k|x_{k-1}, u_k)p(u_k)\tilde{p}(y_k)
\tag{47}
$$

The Forney-factor graph for this generative model can be seen in Figure 18. The precise distributions defined for each component as:

$$
\begin{aligned}
p(y_k|x_k) &= N(x_k, \ \Sigma_y) \\
p(x_k|x_{k-1}, u_k) &= N(h(u_k) + g(x_{k-1}), \ \Sigma_x) \\
p(u_k) &= N(0, \ \Sigma_u)
\end{aligned}
\tag{48}
$$

The component we have neglected to mention above is the preference distribution. The model initialises the preference distribution as vague for the first $T$ time-steps. After which it is restricted to a concentrated distribution around the goal observation:

$$
\tilde{p}(y_k) = \begin{cases} N\left((0,0), \ 10^4 \times \mathbb{I}_2\right), & \text{for } k < T \\ N\left((0.75, 0), \ 10^{-4} \times \mathbb{I}_2\right), & \text{otherwise} \end{cases}
\tag{49}
$$

The algorithm for each time-step using this model can be summarised as comprising of three distinct steps:

**1. Act-Execute-Observce**  - We first select the **mean** of the marginal distribution over the controls, $u_k$. We then execute this actions on the environment and observe the new observations generated from the environment, $y_k$.

**2. Infer**  - Having executed a new action and generated a new observation at time $k$ we will then clamp these within the generative model and perform message passing to infer the posterior distribution of the generative model. This is the step where we require belief propagation.

**3. Slide**   - In this step we instantiate the generative model by specifying the new prior to time $t+1$ as well sliding the inference horizon $T$. Specifically:

More specifically:

$$p_{t+1}(x,y,u) \propto \int \overbrace{p\left(x_{t-1}\right) p\left(x_t | x_{t-1}, u_t\right) p\left(y_t | x_t\right)}^{\text{first slice}} \overbrace{\delta\left(u_t - \widehat{u_t}\right) \delta\left(y_t - \widehat{y_t}\right)}^{\text{action and observation}} dx_{t-1} du_t dy_t$$

<div style="text-align:center">new prior - $p(x_t)$</div>

$$\underbrace{\left( \prod_{k=t+1}^{t+T} p\left(x_k | x_k\right) p\left(x_k | x_{k-1}, u_k\right) p\left(u_k\right) \tilde{p}\left(y_k\right) \right)}_{\text{unaltered mid-section slices}} \tag{50}$$

$$\underbrace{p\left(y_{t+T+1} | x_{t+T+1}\right) p\left(x_{t+T+1} | x_{t+T}, u_{t+T+1}\right) p\left(u_{t+T+1}\right) \tilde{p}\left(y_{t+T+1}\right)}_{\text{add slice at horizon}}$$

Upon inspection we can see that the new prior is nothing but the message coming from the equality node in Figure 18.

**Non-Linearity**   The physical dynamics of the environment are not linear as described in the previous section. The issue with these non-linearities is that incoming messages from an exponential family of distributions is not guaranteed to be part of the exponential family when output. (van de Laar and de Vries, 2019) Thus, we will perform a linear approximation of the change in hidden states over time as given by $F_g$ and $F_f$, this function is denoted as $g$. As well as a linear approximation of the control $F_a$ which is denoted $h$. The exact approximation technique was derived specifically for non-linearities in factor graphs (Petersen et al., 2018).
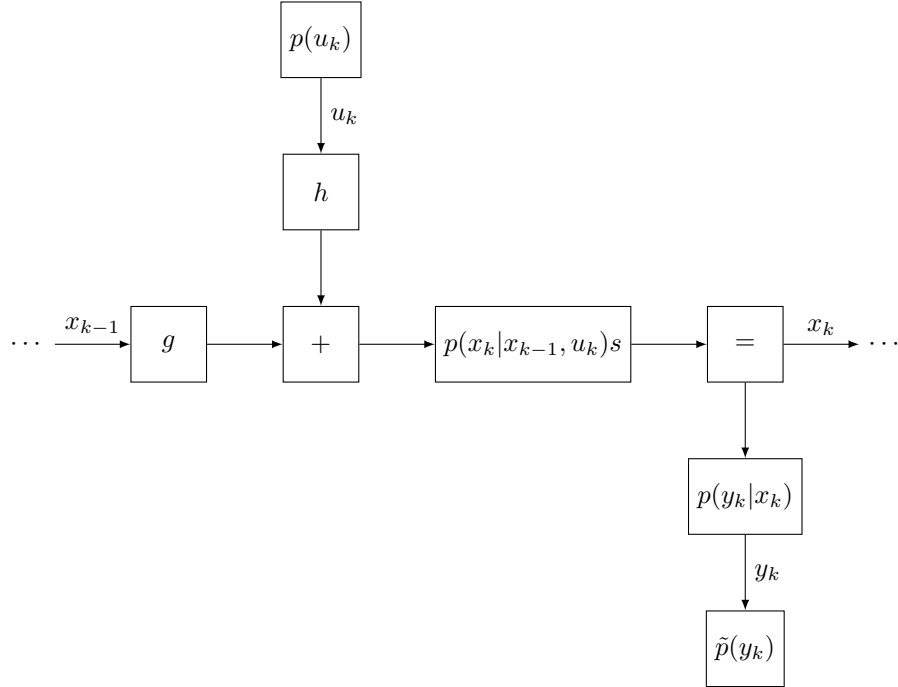


Figure 18: A Forney-factor Graph for the generative model used in the mountain car environment. Each component will be repeated for each time along the time horizon $T$. The factors within the graph that denote the probability distributions have their arguments included in label. This is redundant since the edges dictate the arguments. This was done for clear reference to the generative model in (47).

#### 4.3.4   Results and Discussion

All results in this environment were computed using the a package for Bayesian inference using message passing called RxInfer.jl (Bagaev et al., 2023).
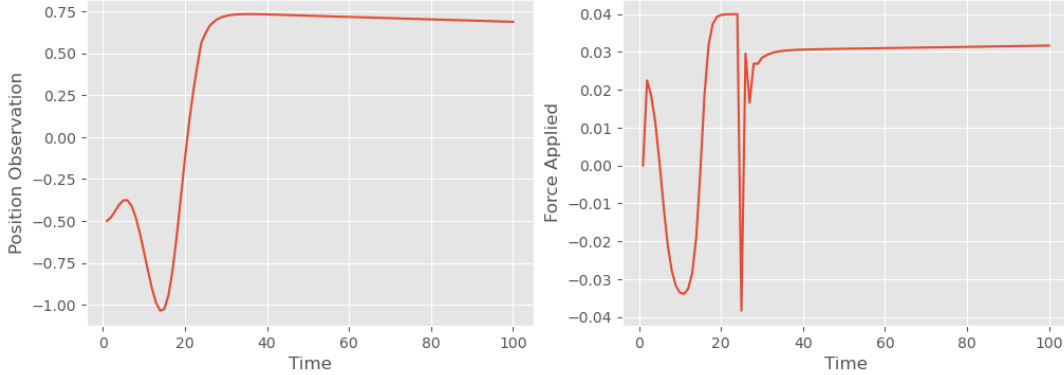


Figure 19: Results from 100 time steps of the active inference agent experiment in the Mountain Car Environment. **Left**: This is a plot of the observations over the 100 time-steps. We see that the agent is able to reach the goal of a height of 0.75. **Right**: This is a plot of the force exerted by the agent over the course of the experiment. We observe that indeed the agent is able to move backwards to gain enough potential energy to reach its target height.

The results depicted in Figure 19 show that the active inference agent successfully reached its target height of 0.75 in the Mountain Car environment. To solve this challenge, the agent had to first move backward to gain enough potential energy, then use its forward momentum to climb high enough on the other side of the valley. This is evident in the force-over-time plot. Notably, the agent demonstrates awareness that gradually approaching the goal is not sufficient, and intelligently applies full forward force, followed by full backward force in a braking-like manoeuvrea highly adaptive behaviour within this environment.

Some degree of exploration is clearly necessary within the environment. If the agent behaves greedily, applying full force directly towards the target, it fails to solve the problem. This is a limitation of the current model. By adjusting the variance of the preference distribution $\tilde{p}$ over time, we explicitly induce exploratory behaviour. In fact, if the agent is set with a strict preference distribution for the entire episode, it acts purely greedily and fails to solve the problem.

In theory, active inference provides an implicit solution to the exploration vs. exploitation dilemma. However, the current model does not compute actions based on the expected free energy (EFE) of each action; instead, actions are sampled from the mean of the marginal distribution. Modifying the model to incorporate this implicit exploration-exploitation tradeoff would be an intriguing direction for further research.

## 5   Conclusion

In this report we have introduced active inference from two distinct conceptual foundations, what we have called the "low road" and "high road" to active inference. The reason for doing this is to shed light on the fact that such a broad reaching framework can be difficult to approach. This split in perspective aims to communicate that there are many ways to start thinking and implementing active inference. The framework can be approached from the broad perspective of a persistent biological system, to a lower level approach through theories of inference in the brain. Each of these approach are distinct the their assumptions and fields but can ultimately be seen as performing active inference.

Understandnig the conceptual basis of the framework is important but it is equally important to put this theory into practice by implementing some of tenets that active inference proposes into models to solve control and planning problems. We showed that a simple model of perception, learning and planning in the form of predictive coding generative model is able to solve the the problem of observing its preferential temperature within an environment by leveraging EFE to plan its actions. Similarly, we were able to demonstrate that a predictive coding model that aims to leverage the correlations between observations through time to balance a pole on the back of a cart. Moreover, we were able to show that belief propogation on a FFG was able to effectively drive a mountain car to a place on a mountain by reversing to create enough potential energy.

These problems are inherently in very simple domains, was have faster and better algorithms to solve them. That was not the point of the applications. The point of the application is that we can take this very grounded mathematical theory to solve problems with very simple and biologically plausible models.

# References

Aggelopoulos, N. C. (2015). Perceptual inference. *Neuroscience & Biobehavioral Reviews*, 55:375–392.

Bagaev, D. and de Vries, B. (2021). Reactive message passing for scalable bayesian inference.

Bagaev, D., Podusenko, A., and de Vries, B. (2023). Rxinfer: A julia package for reactive real-time bayesian inference. *Journal of Open Source Software*, 8(84):5161.

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of mathematical psychology*, 76:198–211.

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

Cox, M., van de Laar, T., and de Vries, B. (2019). A factor graph approach to automated design of bayesian signal processing algorithms. *International Journal of Approximate Reasoning*, 104:185–204.

Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., and Friston, K. (2020). Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology*, 99:102447.

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., and Friston, K. (2008). The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLOS Computational Biology*, 4(8):1–35.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22.

Doya, K., Ishii, S., Pouget, A., and Rao, R. P. (2007). *Bayesian Brain: Probabilistic Approaches to Neural Coding*. The MIT Press.

Florian, R. V. (2005). Correct equations for the dynamics of the cart-pole system.

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, 16(9):1325–1352.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, 360(1456):815–836.

Friston, K. (2008). Hierarchical models in the brain. *PLoS computational biology*, 4(11):e1000211.

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

Friston, K. (2012). A free energy principle for biological systems. *Entropy*, 14(11):2100–2121.

Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *Journal of physiology-Paris*, 100(1-3):70–87.

Friston, K., Klaas, S., Li, B., and Jean, D. (2010). Generalised filtering. *Mathematical Problems in Engineering*, 2010.

Friston, K., Samothrakis, S., and Montague, R. (2012). Active inference and agency: optimal control without cost functions. *Biological cybernetics*, 106:523–541.

Helmholtz, H. v. (1867). Concerning the perceptions in general.

Hodson, R., Mehta, M., and Smith, R. (2023). The empirical status of predictive coding and active inference. *Neuroscience & Biobehavioral Reviews*, page 105473.

Hohwy, J. (2016). The self-evidencing brain. *Nos*, 50(2):259–285.

Ishai, A., Ungerleider, L. G., and Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron*, 28(3):979–990.

Knill, D. C. and Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719.

Miall, R. and Wolpert, D. (1996). Forward models for physiological motor control. *Neural Networks*, 9(8):1265–1279. Four Major Hypotheses in Neuroscience.

Millidge, B. (2019). Combining active inference and hierarchical predictive coding: A tutorial introduction and case study.

Millidge, B. (2021). Applications of the free energy principle to machine learning and neuroscience. *arXiv preprint arXiv:2107.00140*.

Millidge, B., Tang, M., Osanlouy, M., Harper, N. S., and Bogacz, R. (2024). Predictive coding networks for temporal prediction. *PLOS Computational Biology*, 20(4):1–31.

Mishkin, D. and Matas, J. (2016). All you need is a good init.

Moore, A. W. (1990). Efficient memory-based learning for robot control.

Parr, T., Markovic, D., Kiebel, S. J., and Friston, K. J. (2019). Neuronal message passing using mean-field, bethe, and marginal approximations. *Scientific reports*, 9(1):1889.

Parr, T., Pezzulo, G., and Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. The MIT Press.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29(3):241–288.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Penfield, W. (1961). Activation of the record of human experience: Summary of the lister oration delivered at the royal college of surgeons of england* on 27th april 1961. *Annals of the Royal College of Surgeons of England*, 29(2):77.

Petersen, E., Hoffmann, C., and Rostalski, P. (2018). On approximate nonlinear gaussian message passing on factor graphs. In *2018 IEEE Statistical Signal Processing Workshop (SSP)*, page 513517. IEEE.

Podusenko, A., van Erp, B., Koudahl, M., and de Vries, B. (2022). AIDA: An Active Inference-Based Design Agent for Audio Processing Algorithms. *Frontiers in Signal Processing*, 2.

Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87.

Salvatori, T., Song, Y., Yordanov, Y., Millidge, B., Sha, L., Emde, C., Xu, Z., Bogacz, R., and Lukasiewicz, T. (2024). A stable, fast, and fully automatic learning algorithm for predictive coding networks. In *The Twelfth International Conference on Learning Representations*.

Smith, R., Friston, K. J., and Whyte, C. J. (2022). A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, 107:102632.

Song, Y., Lukasiewicz, T., Xu, Z., and Bogacz, R. (2020). Can the brain do backpropagation?—exact implementation of backpropagation in predictive coding networks. *Advances in neural information processing systems*, 33:22566–22579.

Sterpenich, V., Albouy, G., Boly, M., Vandewalle, G., Darsaud, A., Balteau, E., Dang-Vu, T. T., Desseilles, M., D'Argembeau, A., Gais, S., et al. (2007). Sleep-related hippocampo-cortical interplay during emotional memory recollection. *PLoS biology*, 5(11):e282.

Towers, M., Kwiatkowski, A., Terry, J., Balis, J. U., De Cola, G., Deleu, T., Goulão, M., Kallinteris, A., Krimmel, M., KG, A., et al. (2024). Gymnasium: A standard interface for reinforcement learning environments. *arXiv preprint arXiv:2407.17032*.

Trommershauser, J., Maloney, L. T., and Landy, M. S. (2003a). Statistical decision theory and the selection of rapid, goal-directed movements. *JOSA A*, 20(7):1419–1433.

Trommershauser, J., Maloney, L. T., and Landy, M. S. (2003b). Statistical decision theory and trade-offs in the control of motor response. *Spatial vision*, 16(3):255–275.

Ueltzhffer, K. (2018). Deep active inference. *Biological Cybernetics*, 112(6):547573.

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

Welniarz, Q., Worbe, Y., and Gallea, C. (2021). The forward model: A unifying theory for the role of the cerebellum in motor control and sense of agency. *Frontiers in Systems Neuroscience*, 15.

Winn, J., Bishop, C. M., and Jaakkola, T. (2005). Variational message passing. *Journal of Machine Learning Research*, 6(4).

# A Hyperparameters

# B Predictive Coding under the Laplace Approximation

To derive the predictive coding algorithm we assumed the form of the variational posterior distribution to be a Dirac-delta distribution centred around $\phi$. The claim was that that derivation would give use the same result as if we had assumed that the posterior was a Gaussian under the Laplace Approximation, which is the more common approach found in the literature Buckley et al. (2017); Friston (2005). The Laplace approximation defines the variational posterior variance of the distribution to be a function of the mean. More preceisely, we define:

$$q(x|y;\mu) = N(\mu, \sigma(\mu)) \tag{51}$$

O

In this section, we will motivate this claim by deriving the form of the VFE under the Laplace approximation, the goal is to show that this is the exact same form as the VFE when using the Dirac-delta distribution.