entropy

MDPI

Article

On Epistemics in Expected Free Energy for Linear Gaussian State Space Models

Magnus T. Koudahl [1,*], Wouter M. Kouw [1] and Bert de Vries [1,2]

[1] Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands; w.m.kouw@tue.nl (W.M.K.); Bert.de.Vries@tue.nl (B.d.V.)
[2] GN Hearing, JF Kennedylaan 2, 5612 AB Eindhoven, The Netherlands
* Correspondence: m.t.koudahl@tue.nl

Abstract: Active Inference (AIF) is a framework that can be used both to describe information processing in naturally intelligent systems, such as the human brain, and to design synthetic intelligent systems (agents). In this paper we show that Expected Free Energy (EFE) minimisation, a core feature of the framework, does not lead to purposeful explorative behaviour in linear Gaussian dynamical systems. We provide a simple proof that, due to the specific construction used for the EFE, the terms responsible for the exploratory (epistemic) drive become constant in the case of linear Gaussian systems. This renders AIF equivalent to KL control. From a theoretical point of view this is an interesting result since it is generally assumed that EFE minimisation will always introduce an exploratory drive in AIF agents. While the full EFE objective does not lead to exploration in linear Gaussian dynamical systems, the principles of its construction can still be used to design objectives that include an epistemic drive. We provide an in-depth analysis of the mechanics behind the epistemic drive of AIF agents and show how to design objectives for linear Gaussian dynamical systems that do include an epistemic drive. Concretely, we show that focusing solely on epistemics and dispensing with goal-directed terms leads to a form of maximum entropy exploration that is heavily dependent on the type of control signals driving the system. Additive controls do not permit such exploration. From a practical point of view this is an important result since linear Gaussian dynamical systems with additive controls are an extensively used model class, encompassing for instance Linear Quadratic Gaussian controllers. On the other hand, linear Gaussian dynamical systems driven by multiplicative controls such as switching transition matrices do permit an exploratory drive.

Keywords: active inference; epistemics; expected free energy; free energy principle; linear Gaussian dynamical system

Citation: Koudahl, M.T.; Kouw, W.M.; de Vries, B. On Epistemics in Expected Free Energy for Linear Gaussian State Space Models. *Entropy* **2021**, *23*, 1565. https://doi.org/10.3390/e23121565

Academic Editors: Thomas Parr, Manuel Baltieri, Thijs van de Laar, Kai Ueltzhöffer, Daniela Cialfi and Karl Friston

Received: 28 September 2021
Accepted: 23 November 2021
Published: 24 November 2021

1. Introduction

Active Inference (AIF) is a mathematical description of information processing in intelligent systems. In brief it states that agents, originally biological but in later years also synthetic, act to minimise their surprise by seeking out stimuli and states that are compatible with their model of the world. AIF is an attractive framework for designing synthetic agents since AIF agents possess a well-balanced drive towards both explorative (epistemic) and exploitative (pragmatic, goal-driven) behaviour. These characteristics follow from choosing the Expected Free Energy (EFE) as the objective function for planning.

In this paper we explicitly derive the equations for applying AIF in linear Gaussian dynamical systems (LGDS) with the EFE objective. In doing so we uncover a novel result showing that, in the case of linear models, the epistemic term of the EFE objective becomes constant. This means that any application of EFE in LGDS will not lead to exploration and the resulting agents will engage in purely goal-driven behaviour. The proof is given in Section 5.5. The remainder of the paper is structured as an in-depth analysis of the AIF framework and the mechanisms driving its claims to epistemic behaviour. We isolate the

epistemic term of the EFE and identify it as a (bound on) mutual information (MI). We then show that, when considering epistemics in isolation instead of the full EFE construct, it is still possible to generate an epistemic drive using the machinery of AIF. Isolating epistemics corresponds to a special case of EFE where priors on future observations are left unspecified [1,2]. We analyze the behaviour of the resulting epistemic drive and show that, for the case of additive controls, the epistemic drive is independent of state transitions and only depends on the prior variance associated with the belief over the control signal. On the other hand, LGDS driven by multiplicative control signals do exhibit a dependence between state transitions and the epistemic drive. Prior work on AIF in LGDS such as [3–7], have focused mostly on the goal-directed components of the AIF framework. The results, while impressive, largely do not address questions of epistemics and exploration. This means that in cases where AIF is applied to LGDS, EFE and the resulting desirable exploratory drive have so far not been thoroughly investigated. Our results show that, provided the model in question can be cast as a LGDS, incorporating EFE does not lead to meaningful exploration. The present paper makes the following contributions:

- We derive the filtering and planning equations for AIF using EFE in LGDS, Sections 4 and 5.
- We consider the epistemic term of EFE in isolation and show that in the case of additive controls actions become decoupled from state transitions when computing the epistemic term of EFE, Section 5.3. Therefore, we do not find meaningful exploration in this case.
- We show that in the case of multiplicative controls, meaningful exploratory behaviour re-emerges when isolating the epistemic term of EFE, Section 5.4.
- We prove that when considering the full EFE construct, parts of the instrumental and epistemic value terms cancel each other out. This renders the epistemic value constant. In turn, the EFE functional becomes equivalent to KL control plus an additive constant, Section 5.5.
- Finally, we provide simulations that corroborate our claims. We first demonstrate the differences in exploration when considering purely epistemic agents using both additive and multiplicative control signals. Finally we show that LGDS agents optimising the full EFE do not exhibit epistemic drives under any circumstances, Section 6.

The core message is thus that translating AIF to the linear Gaussian case presents unique challenges, specifically because the exploration/exploitation trade-off that follows from EFE minimisation does not manifest. Code to reproduce our experiments is available at github.com/biaslab/efe_lgds (accessed on 19 September 2021).

2. Exploration and Exploitation

In this section we aim to introduce the concepts of exploration and exploitation on intuitive grounds before commencing with our formal analysis. Exploitation refers to goal directed behaviour. An agent that engages in exploitation performs actions that are aimed at optimising some measure of preferences which we will refer to as "Instrumental value". As an example, consider minimisation of mean squared error, cross entropy or a similar cost function. Exploration on the other hand, refers to behaviour directed at collecting information about the environment in which the agent is embedded. An agent that engages in exploration performs actions that are aimed at acquiring further information about its environment. We will refer to any metric that quantifies the value of gathering information as "Epistemic value". Optimising epistemic value biases the agent towards actions that gather information. We will refer to this bias in action selection as an "Epistemic drive". There are many candidates for the epistemic value term. We will briefly consider two that are particularly relevant for the present analysis. This will not be a formal comparison but an intuitive introduction to the qualitative differences in behaviour that can be expected from agents that employ different epistemic value terms. First, we can consider agents that aim to maximise entropy (uncertainty). For such an agent, the epistemic drive biases it towards seeking out areas of state space where uncertainty is high. By repeatedly visiting

uncertain areas of state space, the agent collects observations in said areas which in turn reduces uncertainty. As an example, we can consider an agent trying to navigate an arena. The agent is equipped with a sensor and the arena is subject to strong winds that induce sensor noise by pushing the agent around. In this case maximising entropy drives the agent to seek out parts of the arena where the winds (and corresponding sensor noise) are high. This means the agent collects information primarily in areas where more observations are needed, due to increased sensor noise. Second, we can consider an agent that aims to maximise MI, also known as Information Gain. We provide a formal treatment of MI in Appendix A.4. Intuitively, MI scores the reduction in uncertainty that the agent expects given a particular observation. In the present example, an agent that optimises MI might correctly identify that although windy areas are noisy, collecting information in those areas is unlikely to reduce uncertainty because the winds will remain high. Instead the agent will prefer to move towards areas that have less wind, in order to obtain more accurate measurements. This is the approach taken by AIF agents when optimising EFE [8,9]. Optimising both instrumental and epistemic value terms by selecting actions necessarily entail a trade-off between short term gains (exploitation) and gathering information in order to perform better in the future (exploration). Having agents that are able to optimally balance this trade-off is therefore desirable because it allows for autonomous systems that are able to learn to navigate novel environments in order to achieve desired goals. A core feature of the EFE is that it presents a single objective functional that encompasses both instrumental and epistemic value terms [8,9]. In order to formally unpack how AIF manifests both instrumental and epistemic value terms, we now need to detail the LGDS model that specifies our agent before deriving the equations for computing the EFE objective.

3. Generative Model

AIF is fundamentally a model-based approach [8,10]. As such, the core part of an agent is given by a generative model. Given a generative model, the agent engages in a perception-action loop with its environment. In practice this means the agent will, at any time step, absorb a new observation and emit a new action. The first step is always perception, followed by action selection and emission. Letting $x \in \mathbb{R}^d$ denote observations, $z \in \mathbb{R}^n$ a latent state vector and u actions (we will use "actions" and "controls" interchangeably to refer to u throughout), the generative model for an agent at a single time step, indicated by subscripts t, has the form

$$p(x_t, z_t | u_t, z_{t-1}) = \underbrace{p(x_t|z_t)}_{\text{Likelihood}} \underbrace{p(z_t|z_{t-1}, u_t)}_{\text{State transition}} . \tag{1}$$

This form can be extended, for example by including parameters θ. If applied recursively, this model corresponds to a discrete-time state space model. A common approach when designing AIF agents is to work directly with a policy defined as a *particular* sequence of actions $u_{t+1:T}$ [8,11–13] where the subscript denotes discrete time steps ranging from the next time step $t+1$ to some known planning horizon T. In (1) we indicate this by explicitly conditioning on u_t. This sequence of actions is then considered either as an explicit vector of control signals [8,13–15] or amortised for instance by neural networks [16–19]. Proceeding in this way leads to a particular scheme for action selection which we will detail in Section 5.

In this paper we consider the case of LGDS with multiplicative or additive controls. To clarify the distinction between additive and multiplicative controls, we define "multiplicative controls" as state transitions of the form

$$p(z_t|z_{t-1}, u_t) = \mathcal{N}(z_t|B(u_t)z_{t-1}, \Sigma_z) , \tag{2}$$

where $z_t \in \mathbb{R}^n$ is a latent state vector, u_t is a discrete control signal and $B(u_t)$ is the transition matrix. We consider the case where the control signal functions as a selector variable. Formally we define a vector of candidate transition matrices $[B_1, B_2, \dots B_S]$ and let

$$B(u_t) = \prod_{s=1}^{S} B_s^{u_{ts}} . \tag{3}$$

Here u_t is a one-hot encoded vector $u_t = [u_{t1}, \dots, u_{tS}]$ that takes values in $u_{ts} \in \{0, 1\}$ and where $\sum_{s=1}^{S} u_{ts} = 1$. Each B_s is raised to the power given by u_{ts}, which means that only the selected B_s will be active. The control signal therefore influences state transitions by selecting the transition matrix B directly.

We can visualise this model using the Forney-style Factor Graph (FFG) formalism [20]. In an FFG, each edge represents a variable and each node a factor. An edge is connected to a node if and only if the corresponding variable is an argument of that factor. Each edge connects at most two nodes. When a variable is an argument of more than two factors, we can circumvent the two node per edge limit by linking edges together through equality factors. This effectively creates an auxiliary variable (a new edge) for which the posterior beliefs are constrained to be equal to the beliefs for the original variable. The new edge can also be attached to two factors, so by adding equality factors we can use the same variable as an argument of multiple factors. Observed variables and clamped parameters are denoted by a small black square and selected actions by a small black diamond. The selection mechanism described by (3) is denoted by the multiplexer (MUX) node. Instead of cluttering the graph by drawing the full set of $[B_1, B_2, \cdots B_S]$ candidate transition matrices as separate nodes, we denote them by a shaded circle. The circle contains S nodes and their corresponding outgoing edges all connect to the MUX node. For a further introduction to FFGs, see [21,22]. The FFG of the multiplicative model can be seen in Figure 1.

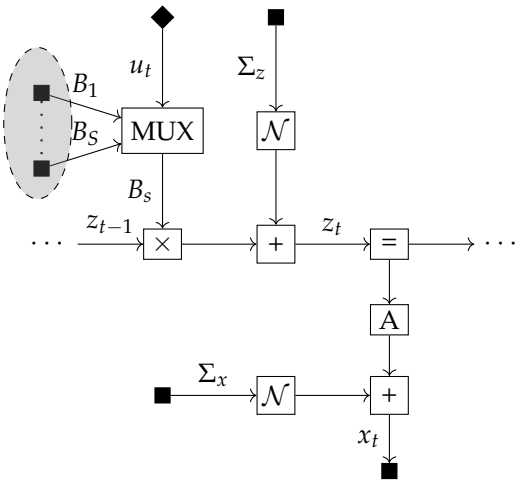

Figure 1. Factor graph of the generative model of an agent with multiplicative control signals.

For comparison, we now consider the case of additive controls. For the additive case the generative model is again given by (1). However exact model specification is a little more involved. We consider transition models of the form

$$p(z_t|z_{t-1}, u_t) = \mathcal{N}(z_t|Bz_{t-1} + b(u_t), \Sigma_z) , \tag{4}$$

where $B \in \mathbb{R}^{n \times n}$ is a known transition matrix and $b(u_t) \in \mathbb{R}^n$ is a vector function that adds to the latent state. To rigorously compare the multiplicative and additive control cases, u_t must remain a categorical selector variable. To that end, we introduce an auxiliary variable b_t. The purpose of b_t is to allow u_t to function as a categorical selector variable. Instead of

selecting between transition matrices B_s, u_t now selects the parameters $\Theta_s = \{\mu_s, \Sigma_s\}$ of a Gaussian input signal. Formally, we will write the generative model as

$$p(z_t|z_{t-1}, u_t) = \int p(z_t|z_{t-1}, b_t)p(b_t|u_t)\mathrm{d}b_t \tag{5a}$$

$$= \int \mathcal{N}(z_t|Bz_{t-1} + b_t, \Sigma_z) \prod_{s=1}^{S} \mathcal{N}(b_t|\Theta_s)^{u_{ts}} \mathrm{d}b_t \,. \tag{5b}$$

where we recognise a similar selection mechanism of (3) in the second term of (5b). Selecting an action means fixing $u_t = \hat{u}_t$ which leads to selecting one of the candidate Gaussian distributions. With only a single Gaussian surviving, integration over b_t becomes straightforward and yields

$$p(z_t|z_{t-1}, \hat{u}_t) = \mathcal{N}(z_t|Bz_{t-1} + \hat{\mu}_{u_t}, \Sigma_z + \hat{\Sigma}_{u_t}) \,, \tag{6}$$

where $\hat{\Theta}_s = \{\hat{\mu}_{u_t}, \hat{\Sigma}_{u_t}\}$ represent the parameters of the Gaussian distribution selected by $\hat{u}_t$.

The factor graph of the additive model is shown in Figure 2 where the MUX node now selects between $\Theta_{1:s}$. In both the multiplicative and additive settings, we employ a likelihood term of the form:

$$p(x_t|z_t) = \mathcal{N}(x_t|Az_t, \Sigma_x) \,, \tag{7}$$

where $A \in \mathbb{R}^{d \times n}$ is a known emission matrix and Σ_x represents measurement or observation noise.

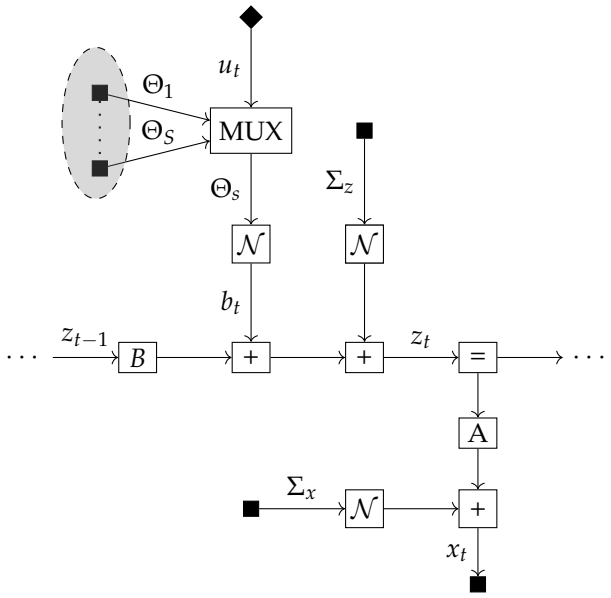

Figure 2. Factor graph of the generative model of an agent with additive controls.

Having established the relevant model structures, we now examine the perception/action loop starting with perception.

4. Perception as Bayesian Filtering

The perception part of the action/perception loop involves making inference about observed data and can be cast as a Bayesian filtering problem. This part of the process describes the agent inferring the hidden state of its environment based on the sequence of actions taken so far, the resulting sequence of states visited and the accompanying

observations. We can write the resulting inference problem in an intuitive way as a prediction-correction process:

$$\underbrace{p(z_t|x_{1:t})}_{\text{posterior}} = \frac{p(x_t|z_t)}{\underbrace{p(x_t|x_{1:t-1})}_{\substack{\text{correction}\\ \text{based on } x_t}}} \times \underbrace{p(z_t|x_{1:t-1})}_{\substack{\text{prediction of } z_t\\ \text{based on } x_{1:t-1}}} . \tag{8}$$

The above shows how the inference over states can be accomplished recursively (due to the model obeying the Markov property) by first computing a prediction for the next hidden state z_t to generate a prior belief which is then updated in a correction step based on the observed data point x_t.

The Bayesian filtering problem is generic. To see how it translates to our case, we can expand the prior predictive in terms of our generative model

$$\underbrace{p(z_t|x_{1:t})}_{\text{state posterior}} = \frac{\overbrace{p(x_t|z_t)}^{\text{likelihood}}}{\underbrace{p(x_t|x_{1:t-1})}_{\text{evidence}}} \overbrace{\iint \underbrace{p(z_t|z_{t-1},u_t)}_{\text{state transition}} \underbrace{\delta(u_t-\hat{u}_t)}_{\substack{\text{control}\\ \text{signal}}} \underbrace{p(z_{t-1}|x_{1:t-1})}_{\text{state prior}} \mathrm{d}z_{t-1}\mathrm{d}u_t}^{\text{prior predictive } p(z_t|x_{1:t-1})} , \tag{9}$$

We use $\delta(u_t - \hat{u}_t)$ where δ is the Dirac-δ to fix the value of u_t to the selected action $\hat{u}_t$ for that particular time step. The particular value chosen for $\hat{u}_t$ is the result of the action selection procedure described in Section 5. The evidence term is given by

$$p(x_t|x_{1:t-1}) = \int p(x_t|z_t)\left(\iint p(z_t|z_{t-1},u_t)\delta(u_t-\hat{u}_t)p(z_{t-1}|x_{1:t-1})\mathrm{d}z_{t-1}\mathrm{d}u_t\right)\mathrm{d}z_t . \tag{10}$$

For the LGDS models considered in this paper, filtering can be performed using the Kalman filtering equations. We will first work this out explicitly in the multiplicative case and then in the additive. To show how to perform filtering in the multiplicative model, we start by assuming that the agent has selected an action $u_t = \hat{u}_t$ by the procedure described in Section 5. We can then calculate the prior predictive distribution of (9) according to our model specification (1) as

$$\begin{aligned}
p(z_t|x_{1:t-1}) &= \iint p(z_t|z_{t-1},u_t)\delta(u_t-\hat{u}_t)p(z_{t-1}|x_{1:t-1})\mathrm{d}z_{t-1}\mathrm{d}u_t && \text{(11a)}\\
&= \iint \underbrace{\mathcal{N}(z_t|B(u_t)z_{t-1},\Sigma_z)}_{\text{State transition}} \underbrace{\delta(u_t-\hat{u}_t)}_{\substack{\text{Selected}\\ \text{control signal}}} \underbrace{\mathcal{N}(z_{t-1}|\mu_{z_{t-1}},\Sigma_{z_{t-1}})}_{\text{State prior}} \mathrm{d}z_{t-1}\mathrm{d}u_t && \text{(11b)}\\
&= \int \mathcal{N}(z_t|B(\hat{u}_t)z_{t-1},\Sigma_z)\mathcal{N}(z_{t-1}|\mu_{z_{t-1}},\Sigma_{z_{t-1}})\mathrm{d}z_{t-1} && \text{(11c)}\\
&= \mathcal{N}(z_t|\underbrace{\hat{B}_t\mu_{z_{t-1}}}_{\mu^-_{z_t}},\underbrace{\hat{B}_t\Sigma_{z_{t-1}}\hat{B}_t^T+\Sigma_z}_{\Sigma^-_{z_t}}) , && \text{(11d)}
\end{aligned}$$

which we recognise as the prediction step of a Kalman filter [23]. We use the superscript $^-$ notation to indicate that the variable in question is not based on the full data set $x_{1:t}$ but instead on a smaller data set $x_{1:t-1}$. In moving from (11b) to (11c) we rely on the sifting property of the Dirac-δ to substitute the selected value for u_t in (3). Since $B(u_t)$ is a function of u_t and u_t is now fixed to $\hat{u}_t$, we can directly substitute the selected parameterisation by setting $B(u_t) = \hat{B}_t$ where $\hat{B}_t$ denotes the parameterisation given by the selected B_s. This takes us from (11b) to (11c). Finally we can rely on standard results for linearly related and jointly Gaussian variables to go from (11c) to (11d), see for example [23] [Appendix A.1] for details of this move in the context of Gaussian state space models or Appendix A.2 for an abbreviated version. For the additive control case, we can calculate the prior predictive distribution in a similar fashion. Starting from the model definition (1), we can write

$$p(z_t|x_{1:t-1}) = \iiint p(z_t|z_{t-1},b_t)p(b_t|u_t)\delta(u_t-\hat{u}_t)p(z_{t-1}|x_{1:t-1})\mathrm{d}u_t\mathrm{d}b_t\mathrm{d}z_{t-1} \tag{12a}$$

$$= \iiint \underbrace{\mathcal{N}(z_t|Bz_{t-1}+b_t,\Sigma_z)\prod_{s=1}^{S}\mathcal{N}(b_t|\mu_s,\Sigma_s)^{u_{ts}}}_{\text{State transition}}\underbrace{\delta(u_t-\hat{u}_t)}_{\substack{\text{Selected}\\ \text{control signal}}}\underbrace{\mathcal{N}(z_{t-1}|\mu_{z_{t-1}},\Sigma_{z_{t-1}})}_{\text{State prior}}\mathrm{d}u_t\mathrm{d}b_t\mathrm{d}z_{t-1} \tag{12b}$$

$$= \iint \mathcal{N}(z_t|Bz_{t-1}+b_t,\Sigma_z)\prod_{s=1}^{S}\mathcal{N}(b_t|\mu_s,\Sigma_s)^{\hat{u}_{ts}}\mathcal{N}(z_{t-1}|\mu_{z_{t-1}},\Sigma_{z_{t-1}})\mathrm{d}b_t\mathrm{d}z_{t-1} \tag{12c}$$

$$= \iint \mathcal{N}(z_t|Bz_{t-1}+b_t,\Sigma_z)\underbrace{\mathcal{N}(b_t|\hat{\mu}_{u_t},\hat{\Sigma}_{u_t})}_{\substack{\text{Selected}\\ \text{parameters, }\hat{\Theta}_s}}\mathcal{N}(z_{t-1}|\mu_{z_{t-1}},\Sigma_{z_{t-1}})\mathrm{d}b_t\mathrm{d}z_{t-1} \tag{12d}$$

$$= \int \mathcal{N}(z_t|Bz_{t-1}+\hat{\mu}_{u_t},\Sigma_z+\hat{\Sigma}_{u_t})\mathcal{N}(z_{t-1}|\mu_{z_{t-1}},\Sigma_{z_{t-1}})\mathrm{d}z_{t-1} \tag{12e}$$

$$= \mathcal{N}(z_t|\underbrace{B\mu_{z_{t-1}}+\hat{\mu}_{u_t}}_{\mu^-_{z_t}},\underbrace{\Sigma_z+\hat{\Sigma}_{u_t}+B\Sigma_{z_{t-1}}B^T}_{\Sigma^-_{z_t}}). \tag{12f}$$

We again denote the selected parameters for the additive control at the k-th time step as $\hat{\mu}_{u_k}$ and $\hat{\Sigma}_{u_k}$. To proceed from (12b) to (12c) we rely on the sifting property of the Dirac-δ and substitute the selected value for u_t. The move from (12c) to (12d) acknowledges that only the selected parameterisation is active once we substitute $u_t = \hat{u}_t$ as covered in Section 3. The final steps from (12d) to (12e) and from (12e) to (12f) uses standard results for multiplication and marginalisation of jointly Gaussian variables for which we again refer to [23] [Appendix A.1] and Appendix A.2. In summary, we see that when $u_t = \hat{u}_t$ the model specification given in (5b) reduces to a standard LGDS and can be updated using the prediction step of the Kalman filtering equations.

Both the additive and multiplicative models use similar likelihood models, meaning they can be updated using the same Kalman correction step. To perform the Kalman correction step, we need to apply Bayes rule

$$\underbrace{p(z_t|x_{1:t})}_{\text{posterior}}\underbrace{p(x_t|x_{1:t-1})}_{\text{evidence}} = \underbrace{p(x_t|z_t)}_{\text{likelihood}}\underbrace{p(z_t|x_{1:t-1})}_{\text{prior}}, \tag{13}$$

where the factor on the right-hand site (RHS) are given and the terms on the left-hand side are the desired factors. This equation can be solved analytically. First, we evaluate the RHS as

$$p(x_t|z_t)p(z_t|x_{1:t-1}) = \mathcal{N}(x_t|Az_t,\Sigma_x)\mathcal{N}(z_t|\mu^-_{z_t},\Sigma^-_{z_t}) \tag{14a}$$

$$= \mathcal{N}\left(\begin{bmatrix}z_t\\ x_t\end{bmatrix}\middle|\begin{bmatrix}\mu^-_{z_t}\\ A\mu^-_{z_t}\end{bmatrix},\begin{bmatrix}\Sigma^-_{z_t} & \Sigma^-_{z_t}A^T\\ A\Sigma^-_{z_t} & A\Sigma^-_{z_t}A^T+\Sigma_x\end{bmatrix}\right). \tag{14b}$$

Then if, for notational convenience, we rewrite the covariance matrix as

$$\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix} \triangleq \begin{bmatrix}\Sigma^-_{z_t} & \Sigma^-_{z_t}A^T\\ A\Sigma^-_{z_t} & A\Sigma^-_{z_t}A^T+\Sigma_x\end{bmatrix}, \tag{15}$$

(14b) can be written as the product of the state posterior

$$p(z_t|x_{1:t}) = \mathcal{N}(z_t|\mu^-_{z_t}+\Sigma_{12}\Sigma_{22}^{-1}(x_t-A\mu^-_{z_t}),\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}), \tag{16}$$

and evidence

$$p(x_t|x_{1:t-1}) = \mathcal{N}(x_t|A\mu^-_{z_t},\Sigma_{22}). \tag{17}$$

due to the theorem for decomposing a multi-variate Gaussian into the product of a conditional distribution [23] [Appendix A.1].

Finally, we can also calculate the conditional distribution $p(x_t|z_t, x_{1:t-1})$ [23] [Appendix A.1]. While this is not required for solving the Bayesian filtering problem, it will prove useful for deriving the epistemic value term (as derived in Appendix A.4) used for the action selection procedure described in Section 5. We can find it as

$$p(x_t|z_t, x_{1:t-1}) = \mathcal{N}(x_t|A\mu_{z_t}^- + \Sigma_{21}\Sigma_{11}^{-1}(z_t - \mu_{z_t}^-), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\,. \tag{18}$$

Having described perception as Bayesian filtering, we now turn our attention to action selection under AIF.

5. Action Selection under Active Inference

When we are interested in constructing AIF agents, arguably the core task is that of action selection. Under AIF we solve this task by first computing a prior over future control signals. Technically, we seek to compute

$$p(u_{t+1:T}) \propto \exp(-G(u_{t+1:T}))\,, \tag{19}$$

i.e., the prior on controls is a softmax function of $G(u_{t+1:T})$ [8] [Equation (7)]. Here, $G(u_{t+1:T})$ denotes the expected free energy (EFE) for a policy that extends into the future until a known horizon T. We further discuss the computation of $G(u_{t+1:T})$ in Section 5.1. To obtain the control prior at time $t+1$, we can marginalise this distribution as

$$p(u_{t+1}) = \int \cdots \int p(u_{t+1:T}) \mathrm{d}u_{t+2} \cdots \mathrm{d}u_T \tag{20a}$$

$$\propto \int \cdots \int \exp(-G(u_{t+1:T})) \mathrm{d}u_{t+2} \cdots \mathrm{d}u_T\,. \tag{20b}$$

If we assume independent control priors for each time step, that is, if we assume $p(u_{t+1:T}) = \prod_{k=t+1}^{T} p(u_k)$, or equivalently,

$$\exp(-G(u_{t+1:T})) = \prod_{k=t+1}^{T} \exp(-G(u_k))\,, \tag{21}$$

then the marginalisation (20) evaluates to

$$p(u_{t+1}) \propto \exp(-G(u_{t+1}))\,. \tag{22}$$

Since the marginalisation procedure is identical for any other time step, we can deduce that the total EFE for a policy is equal to the exponentiated sum of EFEs at individual time steps. That is

$$p(u_{t+1:T}) \propto \exp(-G(u_{t+1:T})) = \prod_{k=t+1}^{T} \exp(-G(u_k)) = \exp\left(-\sum_{k=t+1}^{T} G(u_k)\right). \tag{23}$$

This suggests a recursive scheme over time steps for computing policy priors, similar to the proposal by [13]. In the AIF literature the executed action is then commonly sampled from $p(u_{t+1:T})$ and emitted to the environment [8,13]. Other action selection approaches such as selecting the MAP estimate (the arg max or mode of the distribution) are also possible. We now turn our attention to how the expected free energy is computed.

5.1. Computing G—Expected Free Energy

The EFE is an AIF specific construct that attempts to model what the variational free energy would be at a future time step, conditioned on a particular sequence of actions. Of special interest is the decomposition of EFE into an epistemic (explorative) term and

an instrumental (exploitative) term. It is due to this decomposition that AIF claims an adaptive trade-off between exploration and exploitation [8,13]. Note that we provide the derivation here only for the simplest case. There are extensions to the EFE such as [24] that include additional terms which induce changes in the agent's behaviour. Including these additional terms are not necessary for our core argument so we omit them here and refer interested readers to [24,25]. To show how we arrive at our formulation for EFE, we first need to introduce variational inference. While the filtering equations in Section 4 permit analytical solutions by applying Bayes rule directly, this is often not the case as solving the required integrals can become intractable. In those cases, we can instead approximate the exact solution $p(z_t|x_{1:t})$ by a recognition density $q(z_t)$. Formally we accomplish this by minimising the KL divergence between the exact solution and our recognition density

$$KL[q(z_t)||p(z_t|x_{1:t})] = \int q(z_t) \log \frac{q(z_t)}{p(z_t|x_{1:t})} \mathrm{d}z_t \tag{24}$$

If we now multiply and divide by $p(x_t|x_{1:t-1})$ inside the log-operator,

$$KL[q(z_t)||p(z_t|x_{1:t})] = \int q(z_t) \left[\log \frac{q(z_t)}{p(x_t, z_t|x_{1:t-1})} + \log p(x_t|x_{1:t-1}) \right] \mathrm{d}z_t \tag{25}$$

$$= \underbrace{\int q(z_t) \log \frac{q(z_t)}{p(x_t, z_t|x_{1:t-1})} \mathrm{d}z_t}_{\text{VFE } F[q]} + \underbrace{\log p(x_t|x_{1:t-1})}_{\text{log-evidence}} . \tag{26}$$

We obtain the variational free energy (VFE) $F[q]$ by noting that $\log p(x_t|x_{1:t-1})$ is not dependent on z_t. Since the term $p(x_t, z_t|x_{1:t-1})$ in the denominator of $F[q]$ is given by our generative model, we can choose constraints on $q(z_t)$ to make optimization of (26) tractable. Minimizing $F[q]$ then constitutes an upper bound on $-\log p(x_t|x_{1:t-1})$, meaning we can optimise (26) to obtain an approximate solution to our original inference problem. We define the optimal recognition density q^* as the one that minimises $F[q]$:

$$q^* = \arg\min_q F[q] \tag{27}$$

For further background on variational inference we refer interested readers to the seminal works by [26,27]. Now we are ready to introduce the EFE. To do so, we start by obtaining our best estimate of the time step k in question by integrating out contributions from past time steps as

$$p(x_k, z_k|u_k) = \int p(x_k, z_k|z_{k-1}, u_k) p(z_{k-1}|x_{1:t}) \mathrm{d}z_{k-1} , \tag{28}$$

where we write k instead of t to indicate that we are referring to an arbitrary time step within the planning horizon $t < k \leq T$. $p(z_{k-1}|x_{1:t})$ denotes the posterior state estimate at the previous time step given all available observations. For notational brevity we suppress the dependency on $x_{1:t}$ on the LHS. Unless otherwise noted, all distributions are conditioned on prior observations moving forward. For LGDS $p(z_{k-1}|x_{1:t})$ is available from recursive application of the filtering equations described in Section 4. For $k = t + 1$ it is given by (16) and for $k > t + 1$, $p(z_{k-1}|x_{1:t})$ is given by (11) in the case of multiplicative controls and (12) in the case of additive controls. When $p(z_{k-1}|x_{1:t})$ can not be obtained through application of Bayes rule (providing the exact solution), one can employ variational inference (providing an approximate solution). In that case derivations must instead proceed in terms of the approximate posterior $q(z_{k-1}|x_{1:t})$. Now we can write out the variational free energy conditioned on a particular action $u_k = \hat{u}_k$ and recognition density as $F[q; u_k]$. Note that while $F[q]$ is a *functional* (a function of a function) of q, we also explicitly include conditioning on action given by the *parameter* u_k. To differentiate, we separate them with a semicolon when writing $F[q; u_k]$. Given the factorisation in (23), it

is sufficient to consider a single time step k since we can substitute any value for k. This gives us

$$F[q; u_k] = \int q(z_k|u_k) \log \frac{q(z_k|u_k)}{p(x_k, z_k|u_k)} \mathrm{d}z_k \,. \tag{29}$$

However this expression includes observations x_k which are not available, since we are working with time steps in the future ($t < k \leq T$) and the future is by definition not observed yet. To alleviate this issue, we can take the expectation of this expression with respect to the data generating distribution over observations. When the data generating distribution is available from the generative model, we can equivalently write $p(x_k|z_k)$ instead of $q(x_k|z_k)$. This gives the expression for the expected free energy at the k'th time step:

$$G[q; u_k] = \underbrace{\iint q(x_k|z_k) \overbrace{\left[q(z_k|u_k) \log \frac{q(z_k|u_k)}{p(x_k, z_k|u_k)} \mathrm{d}z_k \right]}^{F[q;u_k] \text{ if } x_k \text{ was observed}} \mathrm{d}x_k}_{\text{Expected } F[q;u_k] \text{ since } x_k \text{ is not yet observed}} \,. \tag{30}$$

As with the VFE in (26), we are interested in the minimum of (30) which once again entails finding q^*. For clarity of notation we define the solution as

$$G(u_k) = G[q^*; u_k] = \arg\min_q G[q; u_k] \,, \tag{31}$$

where $G(u_k)$ is used to compute the policy prior by plugging into (23). Note that $G(u_k)$ is a scalar value that denotes the expectation of $F[q^*; u_k]$ under a particular set of constraints on q and given a specific action u_k. To get an intuition for $G(u_k)$ it can be useful to think of the computation as a two-step procedure consisting of an inner and an outer loop. The inner loop performs variational inference and finds q^* conditioned on an action u_k. The outer loop then computes the resulting EFE by taking the expectation of $F[q^*; u_k]$ under the matching data generating distribution. A core property of EFE is that it introduces an epistemic value term into the optimisation. This leads agents that optimise EFE to seek out areas of state space that have high information gain under the current model, allowing for a principled trade-off between exploration and exploitation [9,28]. To show how this comes about, we can decompose the EFE into a cross-entropy loss and a mutual information (MI) term where the latter quantifies the information gain (in nats or bits) about hidden states z_k from observing outcomes x_k. For the following derivation we will need a bound, the details of which can be found in Appendix A.3. Starting from (30), we can factorise the denominator as $p(x_k, z_k|u_k) = p(z_k|x_k, u_k)p(x_k)$, leading to

$$G(u_k) = \iint q(x_k|z_k) \left[q(z_k|u_k) \log \frac{q(z_k|u_k)}{p(z_k|x_k, u_k)p(x_k)} \mathrm{d}z_k \right] \mathrm{d}x_k \tag{32a}$$

$$= \iint q(x_k|z_k) \left[q(z_k|u_k) \left[\log \frac{q(z_k|u_k)}{p(z_k|x_k, u_k)} - \log p(x_k) \right] \mathrm{d}z_k \right] \mathrm{d}x_k \,. \tag{32b}$$

Now we apply the bound from Appendix A.3 to swap q for p in the denominator. Making use of this inequality is a standard move across the active inference literature [8,11,13,25,29]. The bound becomes exact when we perform exact inference which is the case in the models we consider here and the discrete models often employed in AIF research, see for instance [8,25]). Instead of applying the bound, another option is to utilise (30) as is, see [30] for an example. We proceed as

$$\iint q(x_k|z_k)\left[q(z_k|u_k)\left[\log\frac{q(z_k|u_k)}{p(z_k|x_k,u_k)}-\log p(x_k)\right]\mathrm{d}z_k\right]\mathrm{d}x_k \tag{33}$$
$$\geq \iint q(x_k|z_k)\left[q(z_k|u_k)\left[\log\frac{q(z_k|u_k)}{q(z_k|x_k,u_k)}-\log p(x_k)\right]\mathrm{d}z_k\right]\mathrm{d}x_k\,.$$

Finally we split the integral and integrate over z_k to obtain

$$G(u_k) \geq \iint q(x_k,z_k|u_k)\log\frac{1}{p(x_k)}\mathrm{d}z_k\mathrm{d}x_k - \iint q(x_k,z_k|u_k)\log\frac{q(z_k|x_k,u_k)}{q(z_k|u_k)}\mathrm{d}z_k\mathrm{d}x_k \tag{34a}$$
$$= \int q(x_k|u_k)\log\frac{1}{p(x_k)}\mathrm{d}x_k - \iint q(x_k,z_k|u_k)\log\frac{q(z_k|x_k,u_k)}{q(z_k|u_k)}\mathrm{d}z_k\mathrm{d}x_k \tag{34b}$$
$$= \underbrace{\int q(x_k|u_k)\log\frac{1}{p(x_k)}\mathrm{d}x_k}_{\text{cross-entropy}} - \underbrace{\iint q(x_k,z_k|u_k)\log\frac{q(z_k,x_k|u_k)}{q(z_k|u_k)q(x_k|u_k)}\mathrm{d}z_k\mathrm{d}x_k}_{\text{Mutual Information}}\,. \tag{34c}$$

In the last line we multiply and divide by $q(x_k|u_k)$ to make the MI term explicit. Readers familiar with the broader AIF literature such as [8,9,12] might not immediately recognise the form of (34c) as a common decomposition of the EFE. The equivalence between (34c) and the EFE was originally noted in [8] where the move from (34b) to (34c) is done to show the relation between AIF and InfoMax methods. An advantage of writing the EFE as (34c) is that it cleanly shows how the EFE can be viewed as a combination of two well-known and widely established objectives. From (34c), we see that $G(u_k)$ decomposes into a (bound on a) cross-entropy term minus an MI term. Maximising MI is a known way to induce exploration (i.e., information gain about hidden states from observations) in agents and has been employed in multiple settings both within the control theory [31,32] and reinforcement learning literature [17,30,33]. The cross-entropy loss is between a prior $p(x_k)$ and the posterior distribution $q(x_k|u_k)$ over future observations. This allows for interpreting $p(x_k)$ as a target/goal prior [25,34]. It endows the agent with an instrumental value term that elicits goal-directed behaviour from inferred policies.

Taking this view, $G(u_k)$ can be adequately viewed as scoring the behavior resulting from the action u_k as a balancing act between MI-based explorative and cross-entropy-based exploitative terms. We now examine each of these terms separately to understand how they work in the linear Gaussian case before considering them jointly. We begin by focusing on the MI and how it may drive exploration when considered in isolation.

5.2. Mutual Information Computation

Epistemic behaviour in AIF agents can be considered to be driven by minimising negative MI, as shown in (34c). MI is in general defined as

$$I[x,z] = \iint p(x,z)\log\frac{p(x,z)}{p(x)p(z)}\mathrm{d}x\mathrm{d}z = H[z]-H[z|x] = H[x]-H[x|z]\,. \tag{35}$$

Note that we can write (35) in terms of entropies of either x or z. We can do this since MI is symmetric in its arguments. In the LGDS models we consider, we can evaluate the MI component of $G(u_k)$ as

$$I[x_k,z_k] = \frac{1}{2}\log|I+\Sigma_x^{-1}A\Sigma_{z_k}^{-}A^T|\,. \tag{36}$$

The detailed derivation of (36) can be found in Appendix A.4. To facilitate purely epistemic behaviour, AIF agents can optimise this quantity by selecting appropriate control signals. We will therefore use optimisation of MI as the basis from which to investigate purely epistemic behaviour.

5.3. Pure Exploration as a Function of Additive Control Signals

To show the relation between exploration and controls in the additive case, we now need to show how MI depends on the control signal u_k. For clarity of notation we will do the derivation for the case $k > t+1$, i.e. we will write in terms of the prior predictive $p(z_{k-1}|x_{1:k-2}) = \mathcal{N}(z_{k-1}|\mu^-_{z_{k-1}}, \Sigma^-_{z_{k-1}})$ obtained from (12) instead of $p(z_{k-1}|x_{1:k-1}) = \mathcal{N}(z_{k-1}|\mu_{z_{k-1}}, \Sigma_{z_{k-1}})$ from (16). We do so since observations are not available for $k > t$, meaning we can not perform a full filtering step for the prior time step $k-1$ unless $k = t+1$. For the case $k = t+1$, we can perform filtering for the state prior and can therefore substitute in parameters of $p(z_{k-1}|x_{1:k-1})$ where appropriate. We start the derivation by generating a prediction from our model using (12). We can find the relevant joint distribution at the k'th time step by plugging the result into (14b) to obtain

$$p\left(\begin{bmatrix} z_k \\ x_k \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} z_k \\ x_k \end{bmatrix} \middle\| \begin{bmatrix} B\mu^-_{z_{k-1}} + \hat{\mu}_{u_k} \\ A[B\mu^-_{z_{k-1}} + \hat{\mu}_{u_k}] \end{bmatrix}, \begin{bmatrix} \hat{\Sigma}_{u_k} + \Sigma_z + B\Sigma^-_{z_{k-1}}B^T & [\hat{\Sigma}_{u_k} + \Sigma_z + B\Sigma^-_{z_{k-1}}B^T]A^T \\ A[\hat{\Sigma}_{u_k} + \Sigma_z + B\Sigma^-_{z_{k-1}}B^T] & A[\hat{\Sigma}_{u_k} + \Sigma_z + B\Sigma^-_{z_{k-1}}B^T]A^T + \Sigma_x \end{bmatrix}\right), \tag{37}$$

where we see that the control signal contributes an additive term $\hat{\Sigma}_{u_k}$ which is the variance associated with the selected action. Interestingly, this means that if we let $\hat{\Sigma}_{u_k}$ go to 0, the covariance matrix becomes identical to the multiplicative case detailed in Section 5.4. We plug the marginal over states z_k into (36) to get

$$I[x_k, z_k] = \frac{1}{2}\log|I + \Sigma_x^{-1}A\underbrace{\left[B\Sigma^-_{z_{k-1}}B^T + \Sigma_z + \hat{\Sigma}_{u_k}\right]}_{\Sigma_{z_k}}A^T| \tag{38a}$$

$$= \frac{1}{2}\log|I + \underbrace{\Sigma_x^{-1}AB\Sigma^-_{z_{k-1}}B^TA^T}_{\text{Dynamics dependent}} + \underbrace{\Sigma_x^{-1}A\Sigma_zA^T}_{\text{Policy independent}} + \underbrace{\Sigma_x^{-1}A\hat{\Sigma}_{u_k}A^T}_{\text{Policy dependent}}|. \tag{38b}$$

We notice that the MI decomposes into three terms. We label the first "Dynamics dependent" since it depends only on the transition matrix B, observation noise Σ_x and the prior state variance $\Sigma^-_{z_{k-1}}$. The second term is labeled "Policy independent" since it only depends on the observation noise Σ_x and transition noise Σ_z. Note that neither of the first two terms are influenced by the control signal. The last term is the only one to include $\hat{\Sigma}_{u_k}$ and is therefore "Policy dependent". Crucially, the policy dependent term only depends on the variance of the selected control signal $\hat{\Sigma}_{u_k}$ and the observation noise Σ_x. In other words, it is *independent* of the latent state z_k. Since both $\hat{\Sigma}_{u_k}$ and Σ_x are available a priori, we can precompute the effect of a policy on the epistemic value term before receiving any observations. Further the result is also independent of the trajectory taken by the agent. Therefore in the case of additive controls, maximising MI does not produce targeted exploration. This necessitates the use of a different model structure when epistemic behaviour is desired. A similar result to ours was obtained by [35] for the case of linear dynamics with additive controls.

5.4. Pure Exploration as a Function of Multiplicative Control Signals

To show how epistemic behaviour re-emerges as a function of multiplicative control signals, we now need to show how MI depends on the choice of transition matrix $\hat{B}_k$. We again proceed by generating a prediction from our model using (11). Plugging this into (14b) gives us the joint distribution as

$$p\left(\begin{bmatrix} z_k \\ x_k \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} z_k \\ x_k \end{bmatrix} \middle\| \begin{bmatrix} \hat{B}_k\mu^-_{z_{k-1}} \\ A\hat{B}_k\mu^-_{z_{k-1}} \end{bmatrix}, \begin{bmatrix} \Sigma_z + \hat{B}_k\Sigma^-_{z_{k-1}}\hat{B}_k^T & [\Sigma_z + \hat{B}_k\Sigma^-_{z_{k-1}}\hat{B}_k^T]A^T \\ A[\Sigma_z + \hat{B}_k\Sigma^-_{z_{k-1}}\hat{B}_k^T] & A[\Sigma_z + \hat{B}_k\Sigma^-_{z_{k-1}}\hat{B}_k^T]A^T + \Sigma_x \end{bmatrix}\right). \tag{39}$$

Plugging the above into (36) we find that

$$I[x_k, z_k] = \frac{1}{2}\log|I + \Sigma_x^{-1} A \underbrace{\left[\hat{B}_k \Sigma^-_{z_{k-1}} \hat{B}_k^T + \Sigma_z\right]}_{\Sigma_{z_k}} A^T| \tag{40a}$$

$$= \frac{1}{2}\log|I + \underbrace{\Sigma_x^{-1} A \hat{B}_k \Sigma^-_{z_{k-1}} \hat{B}_k^T A^T}_{\text{Policy dependent}} + \underbrace{\Sigma_x^{-1} A \Sigma_z A^T}_{\text{Policy independent}}| \,. \tag{40b}$$

We see that MI now decomposes into two terms. The first term depends on $\hat{B}_k$ and can be controlled by selecting appropriate transition matrices. The second is *independent* of policy as it only involves process Σ_z and observation noise Σ_x. Note that similar terms also appear in the additive case. The difference between the additive and multiplicative cases is that the choice of transition matrix $\hat{B}_k$ is now under the control of the agent. To maximise MI, the agent must therefore select the $\hat{B}_k$ that maximises the entropy of its latent states z_k. Taking this view offers a nice intuitive explanation for the resulting exploratory drive: To gain the most information, we must perform the actions that lead to the most uncertain outcomes as described in Section 2. To learn the most, we must sample where we know the least.

5.5. Instrumental Value and Expected Free Energy

We now turn our attention to the instrumental value term of $G(u_k)$ after which we analyse the full EFE construct. Recall from (34c) that the instrumental value term is a cross-entropy of the form

$$\int q(x_k|u_k) \log \frac{1}{p(x_k)} \mathrm{d}x_k = \int q(x_k|u_k) \log \frac{q(x_k|u_k)}{p(x_k) q(x_k|u_k)} \mathrm{d}x_k = KL[q(x_k|u_k)||p(x_k)] + H[x_k|u_k] \,. \tag{41}$$

In many cases it is not trivial to obtain $q(x_k|u_k)$ due to intractable integrals. However in the LGDS we are considering, which only involve linear Gaussian relations, it has a tractable expression given by (17). The KL divergence between two Gaussian distributions is given by

$$KL[q(x_k|u_k)||p(x_k)] = \frac{1}{2}\left(\log \frac{|\Sigma_p|}{|\Sigma_q|} + n + (\mu_q - \mu_p)^T \Sigma_p^{-1} (\mu_q - \mu_p) + tr[\Sigma_p^{-1}\Sigma_q]\right). \tag{42}$$

We use subscripts $\{p, q\}$ to denote whether a term comes from $p(x_k)$ or $q(x_k|u_k)$ and use $\{\mu, \Sigma\}$ for the parameters of the corresponding distribution. We can now consider both terms of (34c) jointly in the case of LGDS. Taking (35) and (41) together and making the conditioning on u_k in (35) explicit, we see the full objective comes out as

$$G(u_k) \geq \underbrace{KL[q(x_k|u_k)||p(x_k)] + H[x_k|u_k]}_{\text{Instrumental value}} - \underbrace{H[x_k|u_k] + H[x_k|z_k, u_k]}_{\text{Negative MI}} \tag{43a}$$

$$= \underbrace{KL[q(x_k|u_k)||p(x_k)]}_{\text{Risk}} + \underbrace{H[x_k|z_k, u_k]}_{\text{Ambiguity}} , \tag{43b}$$

where we recover the familiar risk and ambiguity terms. In the specific case of LGDS the inequality becomes an equality when we perform exact inference following the equations laid out in Section 4. However note that when combining the instrumental and epistemic terms instead of considering them in isolation, we perform a seemingly innocuous cancellation and remove the entropy $H[x_k|u_k]$ from the equation. Previously $H[x_k|u_k]$ appeared twice since we considered the epistemic and instrumental terms separately. However when considering the full EFE construct, this is no longer necessary and we are left with just the

ambiguity $H[x_k|z_k, u_k]$. Using the entropy expression for a Gaussian distribution, we can write the ambiguity as

$$H[x_k|z_k, u_k] = \frac{1}{2}\Big(n \log 2\pi + \log|\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}| + n\Big) \tag{44}$$

Recalling the form of the joint given in (14b), we can write each block of the covariance matrix out and find

$$H[x_k|z_k, u_k] = \frac{1}{2}\Big(n \log 2\pi + \log|\underbrace{A\Sigma_{z_k}A^T + \Sigma_x}_{\Sigma_{22}} - \underbrace{A\Sigma_{z_k}}_{\Sigma_{21}}\underbrace{\Sigma_{z_k}^{-1}}_{\Sigma_{11}^{-1}}\underbrace{\Sigma_{z_k}A^T}_{\Sigma_{12}}| + n\Big) \tag{45a}$$

$$= \frac{1}{2}\Big(n \log 2\pi + \log|A\Sigma_{z_k}A^T + \Sigma_x - A\Sigma_{z_k}A^T| + n\Big) \tag{45b}$$

$$= \frac{1}{2}(n \log 2\pi + \log|\Sigma_x| + n) \tag{45c}$$

The cancellation that follows from using a cross-entropy term to drive goal-directed behaviour means that we are left with only the conditional entropy to drive exploration. The above derivation shows that this term is constant and only depends on the observation noise variance Σ_x. This proves that EFE minimisation in LGDS does not lead to exploration. In fact, minimising a KL divergence between a predicted and desired state (the risk term) is the objective of KL control [36] or message passing based simulations of AIF that minimise variational free energy [6,37]. We conclude that in the case of LGDS, the EFE objective is equivalent to the objective of KL control plus an additive constant that depends only on the observation noise variance.

6. Experiments

We investigate the proposed agents in three different settings. First, we investigate pure epistemics in the additive case and show that they do not manifest. Second, we investigate pure epistemics in the multiplicative case and confirm that the agent does indeed perform maximum entropy exploration. Finally we provide comparable experiments for full EFE and show that it indeed reduces to a KL divergence plus a constant.

6.1. Pure Epistemics for Additive Controls

In this section we investigate how the epistemic component of EFE behaves in the additive case. In particular we investigate the effects of different transitions on the epistemic value assigned to a policy. For this experiment the transition model is given by (4). We define the state prior as

$$p(z_{t-1}|z_{1:t-1}) = \mathcal{N}\left(z_{t-1}\middle|\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right), \tag{46}$$

and set both transition and observation noise to identity matrices. We allow the agent a single action by setting $T = t + 1$, which will be the case for all experiments. Further we define the transition matrix B, emission matrix A and observation noise Σ_x as

$$A = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}, \; B = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}, \; \Sigma_x = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}. \tag{47}$$

Note that in the additive case, neither matrix has to be time-varying and so we remove the subscripts. We will also use the same parameterization of Σ_x for all experiments. We compare 4 different candidate parameterisations of the control signal

$$\Theta_1 = \left\{ \mu_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}, \Theta_2 = \left\{ \mu_2 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$
$$\Theta_3 = \left\{ \mu_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \right\}, \Theta_4 = \left\{ \mu_4 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}, \Sigma_4 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \right\}. \quad (48)$$

We choose Θ_1 to function as a baseline. For comparison Θ_2 shares the same covariance matrix but offers a higher displacement of the mean. Θ_3 shares the mean parameter with Θ_1 but has higher variance. Finally Θ_4 increases both the mean and variance over Θ_1. According to (38) varying the mean should not affect the epistemic value since it does not enter into the MI computation. On the other hand, we expect higher variance to affect the policy independent term and lead to increased epistemic value. Consequently we hypothesise that Θ_1 and Θ_2 will lead to identical results in terms of epistemics even though they result in very different posterior states. Following the same line of reasoning, we hypothesise that Θ_3 and Θ_4 will lead to identical results. This in turn implies that Θ_1 and Θ_3 will lead to different values even though the displacement is the same and that a similar pattern will hold for Θ_2 and Θ_4. Results are shown in Table 1, rounded to 3 digits.

Table 1. Epistemic value for additive control signals given state transitions.

Transition	−MI
Θ_1	−1.386
Θ_2	−1.386
Θ_3	−1.609
Θ_4	−1.609

We observe that as hypothesised, MI is not affected by the state transition (Θ_1 and Θ_2 show identical values). We do find an effect of changing the variance which is again independent of the mean (Θ_3 and Θ_4 show identical values). This simple experiment confirms our hypotheses given by (38b): Changing the mean of the control signal does not affect the epistemic term. Changing the variance of the control signal does affect the epistemic term. We conclude that when considering purely epistemic value and additive controls, state transitions and exploration are decoupled. Any effect of the control signal on epistemics is only proportional to the variance of the control, can be pre-computed and does not depend on the agent's trajectory.

6.2. Pure Epistemics for Multiplicative Controls

For comparison, we now perform an analogous experiment for the case of multiplicative controls. We define all quantities in the same way as the additive case. The only change we introduce is defining four transition matrices $B_{1:4}$ to replace $\Theta_{1:4}$. The four candidate transitions we consider are

$$B_1 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}, B_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} B_3 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix} B_4 = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}. \quad (49)$$

Following (40), we hypothesise that larger transitions should lead to lower negative MI by virtue of increasing the value of the policy dependent term. We test this hypothesis across four orders of magnitude and show the results in Table 2.

Table 2. Epistemic value for multiplicative control signals given state transitions.

Transition	−MI
B_1	−0.698
B_2	−1.099
B_3	−4.625
B_4	−9.211

We observe that, as hypothesised, negative MI decreases as a function of the size of the state transition. Larger transitions lead to lower negative MI though the exact relationship is nonlinear in the size of the transition.

6.3. Lack of Epistemics for Expected Free Energy

To investigate the behaviour of AIF agents optimising the full EFE construct, we now repeat both the additive and multiplicative experiments but introduce a goal prior $p(x_t)$. We define the state prior and the goal as

$$p(z_{t-1}|x_{1:t-1}) = \mathcal{N}\left(z_{t-1}\middle|\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right), \quad p(x_t) = \mathcal{N}\left(x_t\middle|\begin{bmatrix}3\\3\end{bmatrix}, \begin{bmatrix}3 & 0\\0 & 3\end{bmatrix}\right). \tag{50}$$

Both the multiplicative and the additive agent employ the same emission matrix A. For the multiplicative agent we further define the set of candidate transition matrices $B_{1:4}$

$$A = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}, B_1 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}, B_2 = \begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}, B_3 = \begin{bmatrix}3 & 0\\0 & 3\end{bmatrix}, B_4 = \begin{bmatrix}4 & 0\\0 & 4\end{bmatrix}. \tag{51}$$

Here we choose B_1 as the identity matrix to serve as a baseline. B_2 moves the agent towards the goal but stops short while B_4 overshoots by the same amount. This means that either transition puts the agent at the same distance from the goal but with different variances and hence different values of the policy dependent term. Finally we allow B_3 to move the agent directly to the goal. For the additive case we set the transition matrix $B = B_1$ and consider the set of candidate parameterisations $\Theta_{1:4}$

$$\begin{aligned}\Theta_1 &= \left\{\mu_1 = \begin{bmatrix}0\\0\end{bmatrix}, \Sigma_1 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right\}, \Theta_2 = \left\{\mu_2 = \begin{bmatrix}2\\2\end{bmatrix}, \Sigma_2 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right\}\\ \Theta_3 &= \left\{\mu_3 = \begin{bmatrix}0\\0\end{bmatrix}, \Sigma_3 = \begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}\right\}, \Theta_4 = \left\{\mu_4 = \begin{bmatrix}2\\2\end{bmatrix}, \Sigma_4 = \begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}\right\}.\end{aligned} \tag{52}$$

where we again vary the mean and variance parameters following a similar logic as for the additive experiments. Notably, both Θ_2 and Θ_4 take the agent directly to the goal but with different variances. We first examine results in the multiplicative case, shown in Table 3.

Table 3. EFE for multiplicative controls.

Transition	KL	Ambiguity	G	Instrumental	Epistemic
B_1	1.33	2.84	4.17	5.27	−1.10
B_2	0.64	2.84	3.48	5.27	−1.79
B_3	1.37	2.84	4.21	6.60	−2.40
B_4	3.54	2.84	6.38	9.27	−2.89

Here the first column shows the KL divergence between the posterior predictive distribution over observations $q(x_t|u_t)$ and the goal prior $p(x_t)$ after the corresponding transition. The second column show the additive constant that corresponds to the ambiguity term. The full EFE is displayed in the third column marked G. Finally the last two

columns display the cross-entropy and negative MI terms as Instrumental and Epistemic value respectively. From Table 3 we see that the lowest KL, and consequently lowest G, is obtained when selecting the B_2 transition matrix. Recall that B_2 stopped short of the goal while B_3 placed the agent directly on top of it. However, because controls are multiplicative, B_3 also results in substantially larger variance which is penalised in the KL. To show that KL is indeed the only driving factor, we can examine the second column, containing the Ambiguity term. We see that it is constant since the observation noise is constant. In turn, we find that the EFE (third column, G) can be written as the sum of the KL and Ambiguity columns. For completeness we have also calculated the cross-entropy (Instrumental value) and MI (Epistemic value). Here we observe similar patterns as in the purely exploratory case; larger transitions lead to lower negative MI. This is accurately balanced by the instrumental term though, highlighting an important point: Our result that EFE does not lead to epistemics is only revealed when we consider a particular way of writing the EFE. If we had instead proceeded from the cross-entropy/MI decomposition, the ambiguity constant would not have materialised. We can create a similar table for the additive case, shown in Table 4.

Table 4. EFE for additive controls.

Transition	KL	Ambiguity	G	Instrumental	Epistemic
Θ_1	3.05	2.84	5.88	7.27	−1.39
Θ_2	0.38	2.84	3.22	4.60	−1.39
Θ_3	3.16	2.84	5.99	7.60	−1.61
Θ_4	0.49	2.84	3.33	4.94	−1.61

We observe that the lowest KL and G corresponds to the transition parameterised by Θ_2 as it takes the agent directly to the goal with small variance. What is interesting about Table 4 is the ambiguity column. We obtain the same additive constant as in the multiplicative case which corroborates our results. Even though the dynamics are different and there are substantial differences in both the instrumental and epistemic value terms, the EFE can still be decomposed as a KL and an additive constant that only depends on the observation noise.

7. Discussion

Viewing EFE from the point of view of mutual information and cross entropy allows for isolating the epistemic and instrumental value terms so they can be investigated separately. This angle was originally taken in [8] and used as a method of relating AIF to other frameworks. Recent work [1,9] investigates a similar decomposition in the discrete case to highlight how pure exploration and exploitation manifest. Our results as well as [1,8,9] all explore how the EFE operates in specific model architectures. Additionally [1,2,8] also note the equivalence between the mutual information term and the objective of optimal Bayesian design. While work such as [29,30,38] have investigated this link in the general case, deriving the specific equations for a wider class of model architectures promises to be a fruitful area for further research. In those cases, the approach followed in our analysis presents a straightforward way to derive the form of the EFE objective by first decomposing it into a pair of known objective functions and then deriving the expressions separately.

Because EFE can be written in terms of marginal/conditional distributions over the latent states z, the analysis presented here applies to any model that utilises a linear likelihood. The results do not depend on the transition model, as demonstrated by our experiments showing similar behaviour for EFE minimization using two different transition models. Our results are consequently equally applicable for a large class of transition models such as auto-regressive models, Gaussian process state space models or deep neural networks without additional adaptation provided the observation model remains linear and Gaussian.

On a similar note, a clear limitation of the present work is the strong reliance on linear observation models. We chose to focus on this case since it allows for an analytical expression of the MI term. However, in general MI is a difficult quantity to compute and one often has to rely on approximations. When approximations are involved, the present analysis is not necessarily applicable, since the decoupling of control signals and epistemics is only demonstrated for the linear case.

In special cases, one can also approximate the joint covariance matrix instead of the mutual information - this is the case for extended Kalman filters for example. In these cases, the present analysis can still apply. Investigating different methods for handling non-linearities is an interesting area for future work on AIF in Gaussian state space models (both linear and non-linear), that can prove useful for neuroscientists and engineers alike.

8. Conclusions

In this paper we have shown how to apply AIF in linear Gaussian state space models. We have derived the expressions for EFE in the linear Gaussian case and investigated how the epistemic value terms function. In particular we have shown that in the case of LGDS, EFE reduces to a KL divergence and an additive constant that only depends on observation noise. We therefore conclude that, in the linear Gaussian case, EFE minimization does not lead to epistemic behaviour.

Additionally we have provided an analysis of the epistemic value term considered in isolation, since the cancellation that leads to an absence of epistemic drive for the full EFE is not present when the instrumental term is not included. Our analysis showed that using additive control signals renders the epistemic value term independent of state transitions. This in turn means that any contribution to the epistemic value term is only dependent on the variance associated with the control signal. In other words, it is independent of any observations the agent might receive and any states it may visit, as was previously demonstrated by [35].

Finally we have shown that utilising multiplicative controls, i.e. selecting from a set of candidate transition matrices, circumvents this problem in the purely epistemic case and provides a meaningful interpretation of controls as inducing epistemic behaviour. The resulting setup is reminiscent of the classical Hidden Markov Model that is commonly seen in AIF. Future work can investigate this link by applying recent advances for the discrete case such as [13] to continuous state spaces with multiplicative control signals.

Author Contributions: Conceptualization, M.T.K.; Formal analysis, M.T.K.; Methodology, M.T.K.; Writing—original draft, M.T.K.; Writing—review & editing, M.T.K., W.M.K. and B.d.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Dutch Research Council grant number NWO perspective program P16-25.

Acknowledgments: This work is part of the research programme Efficient Deep Learning with project number P16-25 project 5, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO). The authors also wish to thank the rest of the BIASLab team for helpful discussions throughout the writing of this manuscript and the reviewers for providing helpful and thorough feedback, in particular reviewer 3 who suggested the example used in Section 2.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Derivations

Appendix A.1. Perception as Bayesian Filtering

To derive the filtering equations in Section 4, we need to infer $p(z_t|x_{1:t})$. We consider a model of the form

$$p(x_t, z_t, u_t|z_{t-1}) = p(x_t|z_t)p(z_t|z_{t-1}, u_t)p(u_t)\,. \tag{A1}$$

We can write the inference task as

$$p(z_t|x_{1:t}) = p(z_t|x_t, x_{1:t-1}) \tag{A2a}$$
$$= \frac{1}{p(x_t|x_{1:t-1})} p(x_t, z_t|x_{1:t-1}) \tag{A2b}$$
$$= \frac{1}{p(x_t|x_{1:t-1})} p(x_t|z_t) p(z_t|x_{1:t-1}) \tag{A2c}$$
$$= \frac{1}{p(x_t|x_{1:t-1})} p(x_t|z_t) \iint p(z_t, z_{t-1}, u_t|x_{1:t-1}) \mathrm{d}z_{t-1} \mathrm{d}u_t \tag{A2d}$$
$$= \underbrace{\frac{1}{p(x_t|x_{1:t-1})}}_{\text{Evidence}} \underbrace{p(x_t|z_t)}_{\text{Likelihood}} \iint \underbrace{p(z_t|z_{t-1}, u_t)}_{\text{State transition}} \underbrace{p(u_t)}_{\substack{\text{Control} \\ \text{prior}}} \underbrace{p(z_{t-1}|x_{1:t-1})}_{\text{State prior}} \mathrm{d}z_{t-1} \mathrm{d}u_t \,. \tag{A2e}$$

Now assuming that observations $x_{1:t} = \hat{x}_{1:t}$ are available and the state prior is available from the last time step, we can select a control by setting

$$p(u_t) = \delta(u_t - \hat{u}_t) \,. \tag{A3}$$

The inference problem becomes

$$p(z_t|\hat{x}_{1:t}) = \frac{1}{p(\hat{x}_t|\hat{x}_{1:t-1})} p(\hat{x}_t|z_t) \iint p(z_t|z_{t-1}, u_t) \delta(u_t - \hat{u}_t) p(z_{t-1}|\hat{x}_{1:t-1}) \mathrm{d}z_{t-1} \mathrm{d}u_t \tag{A4a}$$
$$= \frac{1}{p(\hat{x}_t|\hat{x}_{1:t-1})} p(\hat{x}_t|z_t) \int p(z_t|z_{t-1}, \hat{u}_t) p(z_{t-1}|\hat{x}_{1:t-1}) \mathrm{d}z_{t-1} \,. \tag{A4b}$$

The likelihood is then available from the generative model and the state prior is available from the previous time step. We can find the evidence by marginalising out z_t as

$$p(\hat{x}_t|\hat{x}_{1:t-1}) = \int p(\hat{x}_t|z_t) \int p(z_t|z_{t-1}, \hat{u}_t) p(z_{t-1}|\hat{x}_{1:t-1}) \mathrm{d}z_{t-1} \mathrm{d}z_t \,. \tag{A5}$$

Solving these equations amount to performing Bayesian filtering, synonymous with perception.

Appendix A.2. Linearly Related Gaussian Variables

In this section we show how to obtain a posterior marginal, given linearly related and jointly Gaussian variables. We use this result throughout our derivations in Section 4, for instance when moving from (11c) to (11d). Using x and z for generic variables, the goal is to obtain the posterior

$$p(x) = \int p(x|z) p(z) \mathrm{d}z \tag{A6}$$

where

$$p(z) = \mathcal{N}(z|\mu_z, \Sigma_z) \tag{A7a}$$
$$p(x|z) = \mathcal{N}(x|Az + b, \Sigma_x) \tag{A7b}$$

We can view the problem of obtaining $p(x)$ as first applying a linear transform $(Az + b)$ and then adding Gaussian noise with mean 0 and variance Σ_x. We will deal with each of these steps in turn, starting with the linear transform. The posterior mean μ is given by

$$\mu = \mathbb{E}\big[Az + b\big] = A\mathbb{E}[z] + b = A\mu_z + b \tag{A8}$$

where $\mathbb{E}$ denotes the expectation operation. Here we first factor out the terms that do no depend on z out of the expectation and then identify $\mathbb{E}[z] = \mu_z$ as z is Gaussian. To find the

covariance matrix Σ we can proceed by plugging the terms we know into the definition of covariance

$$\Sigma = \mathbb{E}\Big[(x-\mu)(x-\mu)^T\Big] \tag{A9a}$$

$$= \mathbb{E}\left[\big(\underbrace{Az+b}_{x} - \underbrace{A\mu_z - b}_{\mu}\big)\big(\underbrace{Az+b}_{x} - \underbrace{A\mu_z - b}_{\mu}\big)^T\right] \tag{A9b}$$

$$= \mathbb{E}\Big[(Az - A\mu_z)(Az - A\mu_z)^T\Big] \tag{A9c}$$

Now we can factor A out of the expectation

$$= \mathbb{E}\Big[A(z-\mu_z)(z-\mu_z)^T A^T\Big] \tag{A10a}$$

$$= A\underbrace{\mathbb{E}\Big[(z-\mu_z)(z-\mu_z)^T\Big]}_{\Sigma_z}A^T \tag{A10b}$$

$$= A\Sigma_z A^T. \tag{A10c}$$

In the last line we recognise the definition of the prior covariance matrix Σ_z. To obtain our final result we now need to add Gaussian noise with mean 0 and variance Σ_x. We know that if two Gaussian variables are independent, the variance of their sum is the sum of their variances. We can use this to write the final covariance matrix as

$$\Sigma = A\Sigma_z A^T + \Sigma_x \tag{A11}$$

Which gives the result utilised in the main text as

$$p(x) = \mathcal{N}(x|A\mu_z + b, A\Sigma_z A^T + \Sigma_x) \tag{A12}$$

Appendix A.3. Mutual Information Bound

In this section, we examine the result of substituting q for p in (32). We see that the substitution turns (32) into a bound on the expected free energy that becomes exact when we can do exact inference. We can show this by writing

$$\iint q(x_k, z_k|u_k)\log\frac{q(z_k|u_k)}{q(z_k|x_k,u_k)}\mathrm{d}x_k\mathrm{d}z_k$$

$$= \iint q(x_k, z_k|u_k)\log\frac{q(z_k|u_k)}{q(z_k|x_k,u_k)}\frac{p(z_k|x_k,u_k)}{p(z_k|x_k,u_k)}\mathrm{d}x_k\mathrm{d}z_k \tag{A13a}$$

$$= \iint q(x_k, z_k|u_k)\log\frac{q(z_k|u_k)}{p(z_k|x_k,u_k)}\frac{p(z_k|x_k,u_k)}{q(z_k|x_k,u_k)}\mathrm{d}x_k\mathrm{d}z_k \tag{A13b}$$

$$= \iint q(x_k, z_k|u_k)\log\frac{q(z_k|u_k)}{p(z_k|x_k,u_k)}\mathrm{d}x_k\mathrm{d}z_k - \underbrace{\iint q(x_k, z_k|u_k)\log\frac{q(z_k|x_k,u_k)}{p(z_k|x_k,u_k)}\mathrm{d}x_k\mathrm{d}z_k}_{\mathbb{E}_{q(x_k|u_k)}[\mathrm{KL}[q(z_k|x_k,u_k)||p(z_k|x_k,u_k)]]} \tag{A13c}$$

$$\leq \iint q(x_k, z_k|u_k)\log\frac{q(z_k|u_k)}{p(z_k|x_k,u_k)}\mathrm{d}x_k\mathrm{d}z_k, \tag{A13d}$$

where we recognise the upper bound since the expected KL divergence is non-negative. When the expected KL divergence goes to 0, which is the case when we do exact inference, the bound becomes exact. That is the case for all models we consider in this paper since all relations are linear and all distributions Gaussian.

Appendix A.4. Mutual Information Derivation

In this section, we derive the expression for mutual information in detail. We restate the definition given in (35) here as a starting point

$$I[x,z] = \iint p(x,z)\log\frac{p(x,z)}{p(x)p(z)}\mathrm{d}x\mathrm{d}z = H[z] - H[z|x] = H[x] - H[x|z]\,. \tag{A14}$$

Note that we can write (35) in terms of entropies of either x or z since mutual information is symmetric in its arguments.

Recall that (34c) optimises mutual information between expected observations x_k and latent states z_k. This means that we need the marginal and conditional distributions of x_k in order to calculate the requisite entropies. The marginal is given by (17) and the conditional by (18). Since both distributions are Gaussian, their entropies are available in closed form as

$$H[x_k] = \frac{1}{2}(n\log 2\pi + \log|\Sigma_{22}| + n) \tag{A15a}$$

$$H[x_k|z_k] = \frac{1}{2}\Big(n\log 2\pi + \log|\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}| + n\Big)\,, \tag{A15b}$$

where n denotes the dimensionality. Before we proceed, we need the generalised matrix determinant lemma which can be written as

$$|A + UV^T| = |I + V^T A^{-1}U|\,|A|\,, \tag{A16}$$

where I denotes the identity matrix and A is invertible. By setting $V = U = I$ we obtain a form which will be convenient moving forwards

$$|A + UV^T| = |A + I| = |I + I^T A^{-1} I|\,|A| \tag{A17a}$$

$$= |I + A^{-1}|\,|A|\,, \tag{A17b}$$

which also implies

$$|I - A^{-1}| = \frac{|I - A|}{-|A|}\,. \tag{A18}$$

Now we can write the mutual information as

$$I[x_k, z_k] = H[x_k] - H[x_k|z_k] \tag{A19a}$$

$$= \frac{1}{2}\log|\Sigma_{22}| - \frac{1}{2}\log|\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}| \tag{A19b}$$

$$= -\frac{1}{2}\log|\Sigma_{22}|^{-1}|\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}| \tag{A19c}$$

$$= -\frac{1}{2}\log|I - \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}| \tag{A19d}$$

$$= -\frac{1}{2}\log|I - \underbrace{[A\Sigma_{z_k}A^T + \Sigma_x]^{-1}}_{\Sigma_{22}^{-1}}\underbrace{A\Sigma_{z_k}}_{\Sigma_{21}}\underbrace{\Sigma_{z_k}^{-1}}_{\Sigma_{11}^{-1}}\underbrace{\Sigma_{z_k}A^T}_{\Sigma_{12}}| \tag{A19e}$$

$$= -\frac{1}{2}\log|I - [A\Sigma_{z_k}A^T + \Sigma_x]^{-1}A\Sigma_{z_k}A^T| \tag{A19f}$$

$$= -\frac{1}{2}\log|I - [\underbrace{[A\Sigma_{z_k}A^T]^{-1}A\Sigma_{z_k}A^T}_{\text{Evaluates to } I} + [A\Sigma_{z_k}A^T]^{-1}\Sigma_x]^{-1}| \tag{A19g}$$

$$= -\frac{1}{2}\log|I - [I + [A\Sigma_{z_k}A^T]^{-1}\Sigma_x]^{-1}|\,|\,. \tag{A19h}$$

Using (A18) we can rewrite (A19h) as

$$-\frac{1}{2}\log|I-[I+[A\Sigma_{z_k}A^T]^{-1}\Sigma_x]^{-1}| = -\frac{1}{2}\log\left(\frac{|I-\left[I+(A\Sigma_{z_k}A^T)^{-1}\Sigma_x\right]|}{-|I+(A\Sigma_{z_k}A^T)^{-1}\Sigma_x|}\right) \tag{A20a}$$

$$= -\frac{1}{2}\log\left(\frac{-|(A\Sigma_{z_k}A^T)^{-1}\Sigma_x|}{-|I+(A\Sigma_{z_k}A^T)^{-1}\Sigma_x|}\right) \tag{A20b}$$

$$= -\frac{1}{2}\log\left(\frac{|(A\Sigma_{z_k}A^T)^{-1}\Sigma_x|}{|I+(A\Sigma_{z_k}A^T)^{-1}\Sigma_x|}\right). \tag{A20c}$$

Applying (A17a) in the denominator, we now rewrite (A20c) as

$$-\frac{1}{2}\log\left(\frac{|(A\Sigma_{z_k}A^T)^{-1}\Sigma_x|}{|I+(A\Sigma_{z_k}A^T)^{-1}\Sigma_x|}\right) = -\frac{1}{2}\log\left(\frac{|(A\Sigma_{z_k}A^T)^{-1}\Sigma_x|}{|I+[(A\Sigma_zA^T)^{-1}\Sigma_x]^{-1}||(A\Sigma_{z_k}A^T)^{-1}\Sigma_x|}\right) \tag{A21a}$$

$$= -\frac{1}{2}\log\left(\frac{\cancel{|(A\Sigma_{z_k}A^T)^{-1}\Sigma_x|}}{|I+[(A\Sigma_zA^T)^{-1}\Sigma_x]^{-1}|\cancel{|(A\Sigma_{z_k}A^T)^{-1}\Sigma_x|}}\right) \tag{A21b}$$

$$= -\frac{1}{2}\log\left(\frac{1}{|I+[(A\Sigma_zA^T)^{-1}\Sigma_x]^{-1}|}\right) \tag{A21c}$$

$$= \frac{1}{2}\log\left(\left|I+\left[(A\Sigma_zA^T)^{-1}\Sigma_x\right]^{-1}\right|\right) \tag{A21d}$$

$$= \frac{1}{2}\log|I+\Sigma_x^{-1}A\Sigma_{z_k}A^T|, \tag{A21e}$$

which gives the expression found in (36).

References

1. Sajid, N.; Costa, L.D.; Parr, T.; Friston, K. Active inference, Bayesian optimal design, and expected utility. *arXiv* **2021**, arXiv:2110.04074.
2. Friston, K.; FitzGerald, T.; Rigoli, F.; Schwartenbeck, P.; Doherty, J.; Pezzulo, G. Active inference and learning. *Neurosci. Biobehav. Rev.* **2016**, *68*, 862–879. [CrossRef] [PubMed]
3. Baltieri, M.; Buckley, C. PID Control as a Process of Active Inference with Linear Generative Models. *Entropy* **2019**, *21*, 257. [CrossRef]
4. Friston, K.; Ao, P. Free Energy, Value, and Attractors. *Comput. Math. Methods Med.* **2012**, *2012*, 937860. [CrossRef] [PubMed]
5. Buckley, C.L.; Kim, C.S.; McGregor, S.; Seth, A.K. The free energy principle for action and perception: A mathematical review. *J. Math. Psychol.* **2017**, *81*, 55–79. [CrossRef]
6. van de Laar, T.W.; de Vries, B. Simulating Active Inference Processes by Message Passing. *Front. Robot. AI* **2019**, *6*. [CrossRef]
7. Baltieri, M.; Buckley, C.L. An active inference implementation of phototaxis. *arXiv* **2017**, arXiv:1707.01806.
8. Friston, K.; Rigoli, F.; Ognibene, D.; Mathys, C.; Fitzgerald, T.; Pezzulo, G. Active inference and epistemic value. *Cogn. Neurosci.* **2015**, *6*, 187–214. [CrossRef]
9. Sajid, N.; Ball, P.J.; Friston, K.J. Active inference: Demystified and compared. *arXiv* **2020**, arXiv:1909.10863.
10. Ghavamzadeh, M.; Mannor, S.; Pineau, J.; Tamar, A. Bayesian Reinforcement Learning: A Survey. *arXiv* **2016**, arXiv:1609.04436. doi:10.1561/2200000049.
11. Cullen, M.; Davey, B.; Friston, K.J.; Moran, R.J. Active Inference in OpenAI Gym: A Paradigm for Computational Investigations Into Psychiatric Illness. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **2018**, *3*, 809–818. [CrossRef]
12. Parr, T.; Friston, K.J. Generalised free energy and active inference. *Biol. Cybern.* **2019**, *113*, 495–513. [CrossRef] [PubMed]
13. Friston, K.; Da Costa, L.; Hafner, D.; Hesp, C.; Parr, T. Sophisticated Inference. *arXiv* **2020**, arXiv:2006.04120.
14. Fountas, Z.; Sajid, N.; Mediano, P.A.M.; Friston, K. Deep active inference agents using Monte-Carlo methods. *arXiv* **2020**, arXiv:2006.04176.
15. Tschantz, A.; Seth, A.K.; Buckley, C.L. Learning action-oriented models through active inference. *PLoS Comput. Biol.* **2020**, *16*, e1007805. [CrossRef]
16. Tschantz, A.; Millidge, B.; Seth, A.K.; Buckley, C.L. Reinforcement Learning through Active Inference. *arXiv* **2020**, arXiv:2002.12636.
17. Millidge, B. Deep Active Inference as Variational Policy Gradients. *arXiv* **2019**, arXiv:1907.03876.
18. Tschantz, A.; Baltieri, M.; Seth, A.K.; Buckley, C.L. Scaling active inference. *arXiv* **2019**, arXiv:1911.10601.
19. Ueltzhöffer, K. Deep Active Inference. *Biol. Cybern.* **2018**, *112*, 547–573. [CrossRef]

20. Forney, G.D. Codes on graphs: Normal realizations. *IEEE Trans. Inf. Theory* **2001**, *47*, 520–548. [CrossRef]
21. Loeliger, H.A.; Dauwels, J.; Hu, J.; Korl, S.; Ping, L.; Kschischang, F.R. The Factor Graph Approach to Model-Based Signal Processing. *Proc. IEEE* **2007**, *95*, 1295–1322. [CrossRef]
22. Loeliger, H.A. An introduction to factor graphs. *Signal Process. Mag. IEEE* **2004**, *21*, 28–41. [CrossRef]
23. Sarkka, S. *Bayesian Filtering and Smoothing*; Cambridge University Press: Cambridge, UK, 2013. [CrossRef]
24. Schwartenbeck, P.; FitzGerald, T.; Dolan, R.J.; Friston, K. Exploration, novelty, surprise, and free energy minimization. *Front. Psychol.* **2013**, *4*. [CrossRef] [PubMed]
25. Da Costa, L.; Parr, T.; Sajid, N.; Veselic, S.; Neacsu, V.; Friston, K. Active inference on discrete state-spaces: A synthesis. *arXiv* **2020**, arXiv:2001.07203.
26. Şenöz, İ.; van de Laar, T.; Bagaev, D.; de Vries, B. Variational Message Passing and Local Constraint Manipulation in Factor Graphs. *Entropy* **2021**, *23*, 807. [CrossRef]
27. Blei, D.M.; Kucukelbir, A.; McAuliffe, J.D. Variational Inference: A Review for Statisticians. *J. Am. Stat. Assoc.* **2017**, *112*, 859–877. [CrossRef]
28. Schwartenbeck, P.; Passecker, J.; Hauser, T.U.; FitzGerald, T.H.B.; Kronbichler, M.; Friston, K. Computational mechanisms of curiosity and goal-directed exploration. *Neuroscience* **2018**. [CrossRef]
29. Millidge, B.; Tschantz, A.; Buckley, C.L. Whence the Expected Free Energy? *arXiv* **2020**, arXiv:2004.08128.
30. Hafner, D.; Ortega, P.A.; Ba, J.; Parr, T.; Friston, K.; Heess, N. Action and Perception as Divergence Minimization. *arXiv* **2020**, arXiv:2009.01791.
31. Buisson-Fenet, M.; Solowjow, F.; Trimpe, S. Actively learning gaussian process dynamics. In Proceedings of the 2nd Conference on Learning for Dynamics and Control, Online, 11–12 June 2020; pp. 5–15.
32. Bai, S.; Wang, J.; Chen, F.; Englot, B. Information-theoretic exploration with Bayesian optimization. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 1816–1822. ISSN: 2153-0866. [CrossRef]
33. Berseth, G.; Geng, D.; Devin, C.; Finn, C.; Jayaraman, D.; Levine, S. SMiRL: Surprise Minimizing RL in Dynamic Environments. *arXiv* **2019**, arXiv:1912.05510.
34. Friston, K. A free energy principle for a particular physics. *arXiv* **2019**, arXiv:1906.10184.
35. Solopchuk, O. Information Theoretic Approach to Decision Making in Continuous Domains. Ph.D. Thesis, UCL-Université Catholique de Louvain, Ottigny, Belgium, 2021.
36. Kappen, H.J.; Gómez, V.; Opper, M. Optimal control as a graphical model inference problem. *Mach. Learn.* **2012**, *87*, 159–182. [CrossRef]
37. Schwoebel, S.; Kiebel, S.; Markovic, D. Active Inference, Belief Propagation, and the Bethe Approximation. *Neural Comput.* **2018**, *30*, 2530–2567. [CrossRef] [PubMed]
38. Millidge, B.; Tschantz, A.; Seth, A.; Buckley, C. Understanding the Origin of Information-Seeking Exploration in Probabilistic Objectives for Control. *arXiv* **2021**, arXiv:2103.06859.